B111

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number
WO 01/55179 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/US01/02849

(22) International Filing Date: 29 January 2001 (29.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/178,370	27 January 2000 (27.01.2000)	US
60/178,371	27 January 2000 (27.01.2000)	US
60/178,406	27 January 2000 (27.01.2000)	US
60/178,408	27 January 2000 (27.01.2000)	US
60/178,409	27 January 2000 (27.01.2000)	US
60/178,413	27 January 2000 (27.01.2000)	US
60/178,414	27 January 2000 (27.01.2000)	US
60/180,634	7 February 2000 (07.02.2000)	US
60/220,516	24 July 2000 (24.07.2000)	US
60/221,408	28 July 2000 (28.07.2000)	US
60/221,943	31 July 2000 (31.07.2000)	US
60/257,599	21 December 2000 (21.12.2000)	US
60/260,290	8 January 2001 (08.01.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/178,370 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,371 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,406 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,408 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,409 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,413 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/178,414 (CIP)
Filed on	27 January 2000 (27.01.2000)
US	60/180,634 (CIP)
Filed on	7 February 2000 (07.02.2000)
US	60/220,516 (CIP)
Filed on	24 July 2000 (24.07.2000)
US	60/221,408 (CIP)
Filed on	28 July 2000 (28.07.2000)
US	60/221,943 (CIP)
Filed on	31 July 2000 (31.07.2000)
US	Not furnished (CIP)
Filed on	21 December 2000 (21.12.2000)
US	Not furnished (CIP)
Filed on	8 January 2001 (08.01.2001)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PRAYAGA, Sudhirdas, K** [IN/US]; Apt. 18/136, 140 Mill Street, East Haven, CT 06512 (US). **PADIGARU, Muralidhar** [IN/US]; Apt. 6E, 1579 Rhinelander Avenue, Bronx, NY 10561 (US). **SPYTEK, Kimberly, A.** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **LI, Li** [CN/US]; 478 Oak Street, Chesire, CT 06410 (US). **TCHERNEV, Velizar, T.** [BG/US]; Apt. 156, 1216 S.W. Second Ave., Gainsville, FL 30261 (US). **VERNET, Corine, A., M.** [FR/US]; 4830 N.W. 43rd Street P#253, Gainesville, FL 32060 (US). **PEYMAN, John, A.** [US/US]; 336 West Rock Avenue, New Haven, CT 06515 (US). **MACDOUGALL, John** [CA/US]; 17 Russell Street, Hamden, CT 06517 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: The present invention provides novel isolated NOVX polynucleotides and polypeptides encoded by the NOVX polynucleotides. Also provided are the antibodies that immunospecifically bind to a NOVX polypeptide or any derivative, variant, mutant or fragment of the NOVX polypeptide, polynucleotide or antibody. The invention additionally provides methods in which the NOVX polypeptide, polynucleotide and antibody are utilized in the detection and treatment of a broad range of pathological states, as well as to other uses.

WO 01/55179 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME

TECHNICAL FIELD OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom.

BACKGROUND OF THE INVENTION

Within the animal kingdom, odor detection is a universal tool used for social interaction, predation, and reproduction. Chemosensitivity in vertebrates is modulated by bipolar sensory neurons located in the olfactory epithelium, which extend a single, highly arborized dendrite into the mucosa while projecting axons to relay neurons within the olfactory bulb. The many ciliae on the neurons bear odorant (or olfactory) receptors (ORs), which cause depolarization and formation of action potentials upon contact with specific odorants. ORs may also function as axonal guidance molecules, a necessary function as the sensory neurons are normally renewed continuously through adulthood by underlying populations of basal cells.

The mammalian olfactory system is able to distinguish several thousand odorant molecules. Odorant receptors are believed to be encoded by an extremely large subfamily of G protein-coupled receptors. These receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are likely to underlie the recognition and G-protein-mediated transduction of odorant signals and possibly other chemosensing responses as well. The genes encoding these receptors are devoid of introns within their coding regions. Schurmans

and co-workers cloned a member of this family of genes, OLFR1, from a genomic library by cross-hybridization with a gene fragment obtained by PCR. See *Schurmans et al.*, Cytogenet. Cell Genet., 1993, 63(3):200. By isotopic *in situ* hybridization, they mapped the gene to 17p13-p12 with a peak at band 17p13. A minor peak was detected on chromosome 3, with a maximum in the region 3q13-q21. After MspI digestion, a restriction fragment length polymorphism (RFLP) was demonstrated. Using this in a study of 3 CEPH pedigrees, they demonstrated linkage with D17S126 at 17pter-p12; maximum lod = 3.6 at theta = 0.0. Used as a probe on Southern blots under moderately stringent conditions, the cDNA hybridized to at least 3 closely related genes. Ben-Arie and colleagues cloned 16 human OLFR genes, all from 17p13.3. See *Ben-Arie et al.*, Hum. Mol. Genet., 1994, 3(2):229. The intronless coding regions are mapped to a 350-kb contiguous cluster, with an average intergenic separation of 15 kb. The OLFR genes in the cluster belong to 4 different gene subfamilies, displaying as much sequence variability as any randomly selected group of OLFRs. This suggested that the cluster may be one of several copies of an ancestral OLFR gene repertoire whose existence may have predated the divergence of mammals. Localization to 17p13.3 was performed by fluorescence *in situ* hybridization as well as by somatic cell hybrid mapping.

Previously, OR genes cloned in different species were from disparate locations in the respective genomes. The human OR genes, on the other hand, lack introns and may be segregated into four different gene subfamilies, displaying great sequence variability. These genes are primarily expressed in olfactory epithelium, but may be found in other chemoresponsive cells and tissues as well.

Blache and co-workers used polymerase chain reaction (PCR) to clone an intronless cDNA encoding a new member (named OL2) of the G protein-coupled receptor superfamily.See *Blache et al.*, Biochem. Biophys. Res. Commun., 1998, 242(3):669. The coding region of the rat OL2 receptor gene predicts a seven transmembrane domain receptor of 315 amino acids. OL2 has 46.4 percent amino acid identity with OL1, an olfactory receptor expressed in the developing rat heart, and slightly lower percent identities with several other olfactory receptors. PCR analysis reveals that the transcript is present mainly in the rat spleen and in a mouse insulin-secreting cell line (MIN6). No correlation was found between the tissue distribution of OL2 and that of the olfaction-related GTP-binding protein Golf alpha subunit. These findings suggest a

role for this new hypothetical G-protein coupled receptor and for its still unknown ligand in the spleen and in the insulin-secreting beta cells.

Olfactory loss may be induced by trauma or by neoplastic growths in the olfactory neuroepithelium. There is currently no treatment available that effectively restores olfaction in the case of sensorineural olfactory losses. See Harrison's Principles of Internal Medicine, 14th Ed., Fauci, AS *et al.* (eds.), McGraw-Hill, New York, 1998, 173. There thus remains a need for effective treatment to restore olfaction in pathologies related to neural olfactory loss.

SUMMARY OF THE INVENTION

The invention is based, in part, upon the discovery of novel polynucleotide sequences encoding novel polypeptides.

Accordingly, in one aspect, the invention provides an isolated nucleic acid molecule that includes the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 or a fragment, homolog, analog or derivative thereof. The nucleic acid can include, *e.g.*, a nucleic acid sequence encoding a polypeptide at least 85% identical to a polypeptide that includes the amino acid sequences of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36. The nucleic acid can be, *e.g.*, a genomic DNA fragment, or a cDNA molecule.

Also included in the invention is a vector containing one or more of the nucleic acids described herein, and a cell containing the vectors or nucleic acids described herein.

The invention is also directed to host cells transformed with a vector comprising any of the nucleic acid molecules described above.

In another aspect, the invention includes a pharmaceutical composition that includes a NOVX nucleic acid and a pharmaceutically acceptable carrier or diluent.

In a further aspect, the invention includes a substantially purified NOVX polypeptide, *e.g.*, any of the NOVX polypeptides encoded by an NOVX nucleic acid, and fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition that includes an NOVX polypeptide and a pharmaceutically acceptable carrier or diluent.

In still a further aspect, the invention provides an antibody that binds specifically to an NOVX polypeptide. The antibody can be, *e.g.*, a monoclonal or polyclonal antibody, and

fragments, homologs, analogs, and derivatives thereof. The invention also includes a pharmaceutical composition including NOVX antibody and a pharmaceutically acceptable carrier or diluent. The invention is also directed to isolated antibodies that bind to an epitope on a polypeptide encoded by any of the nucleic acid molecules described above.

The invention also includes kits comprising any of the pharmaceutical compositions described above.

The invention further provides a method for producing an NOVX polypeptide by providing a cell containing an NOVX nucleic acid, *e.g.*, a vector that includes an NOVX nucleic acid, and culturing the cell under conditions sufficient to express the NOVX polypeptide encoded by the nucleic acid. The expressed NOVX polypeptide is then recovered from the cell. Preferably, the cell produces little or no endogenous NOVX polypeptide. The cell can be, *e.g.*, a prokaryotic cell or eukaryotic cell.

The invention is also directed to methods of identifying an NOVX polypeptide or nucleic acid in a sample by contacting the sample with a compound that specifically binds to the polypeptide or nucleic acid, and detecting complex formation, if present.

The invention further provides methods of identifying a compound that modulates the activity of an NOVX polypeptide by contacting an NOVX polypeptide with a compound and determining whether the NOVX polypeptide activity is modified.

The invention is also directed to compounds that modulate NOVX polypeptide activity identified by contacting an NOVX polypeptide with the compound and determining whether the compound modifies activity of the NOVX polypeptide, binds to the NOVX polypeptide, or binds to a nucleic acid molecule encoding an NOVX polypeptide.

In another aspect, the invention provides a method of determining the presence of or predisposition of an NOVX-associated disorder in a subject. The method includes providing a sample from the subject and measuring the amount of NOVX polypeptide in the subject sample. The amount of NOVX polypeptide in the subject sample is then compared to the amount of NOVX polypeptide in a control sample. An alteration in the amount of NOVX polypeptide in the subject protein sample relative to the amount of NOVX polypeptide in the control protein sample indicates the subject has a tissue proliferation-associated condition. A control sample is preferably taken from a matched individual, *i.e.*, an individual of similar age, sex, or other general condition but who is not suspected of having a tissue proliferation-associated condition.

Alternatively, the control sample may be taken from the subject at a time when the subject is not suspected of having a tissue proliferation-associated disorder. In some embodiments, the NOVX is detected using an NOVX antibody.

In a further aspect, the invention provides a method of determining the presence of or predisposition of an NOVX-associated disorder in a subject. The method includes providing a nucleic acid sample, *e.g.*, RNA or DNA, or both, from the subject and measuring the amount of the NOVX nucleic acid in the subject nucleic acid sample. The amount of NOVX nucleic acid sample in the subject nucleic acid is then compared to the amount of an NOVX nucleic acid in a control sample. An alteration in the amount of NOVX nucleic acid in the sample relative to the amount of NOVX in the control sample indicates the subject has a NOVX-associated disorder.

In a still further aspect, the invention provides a method of treating or preventing or delaying an NOVX-associated disorder. The method includes administering to a subject in which such treatment or prevention or delay is desired an NOVX nucleic acid, an NOVX polypeptide, or an NOVX antibody in an amount sufficient to treat, prevent, or delay a NOVX-associated disorder in the subject.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

Olfactory receptors (ORs) are the largest family of G-protein-coupled receptors (GPCRs) and belong to the first family (Class A) of GPCRs, along with catecholamine receptors and opsins. The OR family contains over 1,000 members that traverse the phylogenetic spectrum from *C. elegans* to mammals. ORs most likely emerged from prototypic GPCRs several times

independently, extending the structural diversity necessary both within and between species in order to differentiate the multitude of ligands. Individual olfactory sensory neurons are predicted to express a single, or at most a few, ORs. All ORs are believed to contain seven α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. The pocket of OR ligand binding is expected to be between the second and sixth transmembrane domains of the proteins. Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%, and genes greater than 80% identical to one another at the amino acid level are considered to belong to the same subfamily.

Since the first ORs were cloned in 1991, outstanding progress has been made into their mechanisms of action and potential dysregulation during disease and disorder. It is understood that some human diseases result from rare mutations within GPCRs. Drug discovery avenues could be used to produce highly specific compounds on the basis of minute structural differences of OR subtypes, which are now being appreciated with *in vivo* manipulation of OR levels in transgenic and knock-out animals. Furthermore, due to the intracellular homogeneity and ligand specificity of ORs, renewal of specific odorant-sensing neurons lost in disease or disorder is possible by the introduction of individual ORs into basal cells. Additionally, new therapeutic strategies may be elucidated by further study of so-called orphan receptors, whose ligand(s) remain to be discovered.

OR proteins bind odorant ligands and transmit a G-protein-mediated intracellular signal, resulting in generation of an action potential. The accumulation of DNA sequences of hundreds of OR genes provides an opportunity to predict features related to their structure, function and evolutionary diversification. *See* Pilpel Y, et.al., Essays Biochem 1998;33:93-104. The OR repertoire has evolved a variable ligand-binding site that ascertains recognition of multiple odorants, coupled to constant regions that mediate the cAMP-mediated signal transduction. The cellular second messenger underlies the responses to diverse odorants through the direct gating of olfactory-specific cation channels. This situation necessitates a mechanism of cellular exclusion, whereby each sensory neuron expresses only one receptor type, which in turn influences axonal projections. A 'synaptic image' of the OR repertoire thus encodes the detected odorant in the central nervous system.

The ability to distinguish different odors depends on a large number of different odorant receptors (ORs). ORs are expressed by nasal olfactory sensory neurons, and each neuron expresses only 1 allele of a single OR gene. In the nose, different sets of ORs are expressed in distinct spatial zones. Neurons that express the same OR gene are located in the same zone; however, in that zone they are randomly interspersed with neurons expressing other ORs. When the cell chooses an OR gene for expression, it may be restricted to a specific zonal gene set, but it may select from that set by a stochastic mechanism. Proposed models of OR gene choice fall into 2 classes: locus-dependent and locus-independent. Locus-dependent models posit that OR genes are clustered in the genome, perhaps with members of different zonal gene sets clustered at distinct loci. In contrast, locus-independent models do not require that OR genes be clustered.

OR genes have been mapped to 11 different regions on 7 chromosomes. These loci lie within paralogous chromosomal regions that appear to have arisen by duplications of large chromosomal domains followed by extensive gene duplication and divergence. Studies have shown that OR genes expressed in the same zone map to numerous loci; moreover, a single locus can contain genes expressed in different zones. These findings raised the possibility that OR gene choice is locus-independent or involved consecutive stochastic choices.

Issel-Tarver and Rine (1996) characterized 4 members of the canine olfactory receptor gene family. The 4 subfamilies comprised genes expressed exclusively in olfactory epithelium. Analysis of large DNA fragments using Southern blots of pulsed field gels indicated that subfamily members were clustered together, and that two of the subfamilies were closely linked in the dog genome. Analysis of the four olfactory receptor gene subfamilies in 26 breeds of dog provided evidence that the number of genes per subfamily was stable in spite of differential selection on the basis of olfactory acuity in scent hounds, sight hounds, and toy breeds.

Issel-Tarver and Rine (1997) performed a comparative study of four subfamilies of olfactory receptor genes first identified in the dog to assess changes in the gene family during mammalian evolution, and to begin linking the dog genetic map to that of humans. These four families were designated by them OLF1, OLF2, OLF3, and OLF4 in the canine genome. The subfamilies represented by these four genes range in size from 2 to 20 genes. They are all expressed in canine olfactory epithelium but were not detectably expressed in canine lung, liver, ovary, spleen, testis, or tongue. The OLF1 and OLF2 subfamilies are tightly linked in the dog genome and also in the human genome. The smallest family is represented by the canine OLF1

gene. Using dog gene probes individually to hybridize to Southern blots of genomic DNA from 24 somatic cell hybrid lines. They showed that the human homologous OLF1 subfamily maps to human chromosome 11. The human gene with the strongest similarity to the canine OLF2 gene also mapped to chromosome 11. Both members of the human subfamily that hybridized to canine OLF3 were located on chromosome 7. It was difficult to determine to which chromosome or chromosomes the human genes that hybridized to the canine OLF4 probe mapped. This subfamily is large in mouse and hamster as well as human, so the rodent background largely obscured the human cross-hybridizing bands. It was possible, however, to discern some human-specific bands in blots corresponding to human chromosome 19. They refined the mapping of the human OLF1 homolog by hybridization to YACs that map to 11q11. In dogs, the OLF1 and OLF2 subfamilies are within 45 kb of one another (Issel-Tarver and Rine (1996)).

Issel-Tarver and Rine (1997) demonstrated that in the human OLF1 and OLF2 homologs are likewise closely linked. By studying YACs, Issel-Tarver and Rine (1997) found that the human OLF3 homolog maps to 7q35. A chromosome 19-specific cosmid library was screened by hybridization with the canine OLF4 gene probe, and clones that hybridized strongly to the probe even at high stringency were localized to 19p13.1 and 19p13.2. These clones accounted, however, for a small fraction of the homologous human bands.

Rouquier et al. (1998) demonstrated that members of the olfactory receptor gene family are distributed on all but a few human chromosomes. Through fluorescence *in situ* hybridization analysis, they showed that OR sequences reside at more than 25 locations in the human genome. Their distribution was biased for terminal bands of chromosome arms. Flow-sorted chromosomes were used to isolate 87 OR sequences derived from 16 chromosomes. Their sequence relationships indicated the inter- and intrachromosomal duplications responsible for OR family expansion. Rouquier et al. (1998) determined that the human genome has accumulated a striking number of dysfunctional copies: 72% of these sequences were found to be pseudogenes. ORF-containing sequences predominate on chromosomes 7, 16, and 17.

Trask et al. (1998) characterized a subtelomeric DNA duplication that provided insight into the variability, complexity, and evolutionary history of that unusual region of the human genome, the telomere. Using a DNA segment cloned from chromosome 19, they demonstrated that the blocks of DNA sequence shared by different chromosomes can be very large and highly similar. Three chromosomes appeared to have contained the sequence before humans migrated

around the world. In contrast to its multicopy distribution in humans, this subtelomeric block maps predominantly to a single locus in chimpanzee and gorilla, that site being nonorthologous to any of the locations in the human genome. Three new members of the olfactory receptor (OR) gene family were found to be duplicated within this large segment of DNA, which was found to be present at 3q, 15q, and 19p in each of 45 unrelated humans sampled from various populations. From its sequence, one of the OR genes in this duplicated block appeared to be potentially functional. The findings raised the possibility that functional diversity in the OR family is generated in part through duplications and interchromosomal rearrangements of the DNA near human telomeres.

Mombaerts (1999) reviewed the molecular biology of the odorant receptor (OR) genes in vertebrates. Buck and Axel (1991) discovered this large family of genes encoding putative odorant receptor genes. Zhao et al. (1998) provided functional proof that one OR gene encodes a receptor for odorants. The isolation of OR genes from the rat by Buck and Axel (1991) was based on three assumptions. First, ORs are likely G protein-coupled receptors, which characteristically are 7-transmembrane proteins. Second, ORs are likely members of a multigene family of considerable size, because an immense number of chemicals with vastly different structures can be detected and discriminated by the vertebrate olfactory system. Third, ORs are likely expressed selectively in olfactory sensory neurons. Ben-Arie et al. (1994) focused attention on a cluster of human OR genes on 17p, to which the first human OR gene, OR1D2, had been mapped by Schurmans et al. (1993). According to Mombaerts (1999), the sequences of more than 150 human OR clones had been reported.

The human OR genes differ markedly from their counterparts in other species by their high frequency of pseudogenes, except the testicular OR genes. Research showed that individual olfactory sensory neurons express a small subset of the OR repertoire. In rat and mouse, axons of neurons expressing the same OR converge onto defined glomeruli in the olfactory bulb.

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their polypeptides. The sequences are collectively referred to as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated otherwise, "NOVX" is meant to refer to any of the novel sequences

disclosed herein. Table 1 provides a summary of the NOVX nucleic acids and their encoded polypeptides. Example 1 provides a description of how the novel nucleic acids were identified.

NOVX POLYNUCLEOTIDES AND POLYPEPTIDES

OR proteins have seven transmembrane α-helices separated by three extracellular and three cytoplasmic loops, with an extracellular amino-terminus and a cytoplasmic carboxy-terminus. Multiple sequence aligment suggests that the ligand-binding domain of the ORs is between the second and sixth transmembrane domains. Thus, any of NOV1–NOV18 is predicted to have a seven transmembrane region and is similar in that region to a representative GPCR, *e.g.* dopamine (GPCR) (GenBank Accession No.: P20288). Because the OR family of the GPCR superfamily is a group of related proteins specifically located at the ciliated surface of olfactory sensory neurons in the nasal epithelium and are involved in the initial steps of the olfactory signal transduction cascade, any of the polynucleotides and/or polypeptides described below can be used to detect nasal epithelial neuronal tissue.

Based on its relatedness to the known members of the OR family of the GPCR superfamily, any of the polynucleotides and/or polypeptides described below satisfy a need in the art by providing new diagnostic or therapeutic compositions useful in the treatment of disorders associated with alterations in the expression of members of OR family-like proteins. Nucleic acids, polypeptides, antibodies, and other compositions of the present invention are useful in the treatment and/or diagnosis of a variety of diseases and pathologies, including by way of nonlimiting example, those involving neurogenesis, cancer and wound healing.

Overall amino acid sequence identity within the mammalian OR family ranges from 45% to >80%. OR genes that are 80% or more identical to each other at the amino acid level are considered by convention to belong to the same subfamily. See *Dryer and Berghard*, Trends in Pharmacological Sciences, 1999, 20:413.

TABLE 1. Sequences and Corresponding SEQ ID Numbers

ABLE 1. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (polypeptide)	Homology
1	AC019108_B (GPCR2_87840)	1	2	OR GPCR
2	AC019108_C (GPCR3_07361)	3	4	OR GPCR
3	AC019108_D (GPCR4_123201)	5	6	OR GPCR
4	AC019108_E (GPCR5_147441)	7	8	OR GPCR
5	AC019108_F (GPCR1_36560)	9	10	OR GPCR
6	AC019108_G (GPCR7_131681)	11	12	OR GPCR
7	AC019108D	13	14	OR GPCR
8	CG50373-01	15	16	OR GPCR

9	AC019108E_da1	17	18	OR GPCR
10	AC019108_H (GPCR8_3441)	19	20	OR GPCR
11	nh0413n10_A_da2	21	22	OR GPCR
12	AC0170103_B	23	24	OR GPCR
13	nh0413n10_A1	25	26	OR GPCR
14	AC019108_F_da1	27	28	OR GPCR
15	AC019108D_da2	29	30	OR GPCR
16	nh0413n10_A_da4	31	32	OR GPCR
17	CG55604-06	33	34	OR GPCR
18	nh0413n10_A_da3	35	36	OR-GPCR

Where OR GPCR is an odorant receptor of the G-protein coupled-receptor family.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit, *e.g.*, neurogenesis, cell differentiation, cell motility, cell proliferation and angiogenesis. Additional utilities for the NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOV1

A NOV1 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins. A NOV1 nucleic acid and its encoded polypeptide includes the sequences shown in Table 2. The disclosed nucleic acid (SEQ ID NO:1) is 1,039 nucleotides in length and contains an open reading frame (ORF) that begins with an ATG initiation codon at nucleotides 27-29 and ends with a TAA stop codon at nucleotides 963-965. The representative ORF encodes a 312 amino acid polypeptide (SEQ ID NO:2). Putative untranslated regions upstream and downstream of the coding sequence are underlined in SEQ ID NO: 1.

TABLE 2

GTTTGATGTTCCAGGACTAATTTGACATGAAGATAAAGAATCACACTCCAGTAACTGAGGTCCCCCTGATGGGAATCC
CTCATACAAAGGGGATGGAAAATGTGCTTTTTGTCTTATTTCTGGCCTTCTACCTCTTCACCTTGCTGGGGAACCTAC

TCATTCTTCTGGCCGTCCTCACTTTCTCCAACCTCCACACCCCCATGTATTTCTTCCTGGGAAACCTGTCTGTGTTTG
ACATATTTTTCCCTTCAGTGAGTTCCCCCAAAATGATGCTCTGCTTAGTGGGACAAAGCTGCACCATCTCTTTCCAGG
GTTGTGCCTCCCAGCTCTTCTTTCACCATTTCCTGGGTTGCACCGAGTGTTTCCTGTACACTGTGATGGCCTATGACC
GATTTGCAGCCATCTGCCACCCTTTGCCATACACGGTCATCATGAAACGCAGGGTGTGTGCCCTCCTGACGCTAGGCA
CCTGGACGGGGAGCTGTCTGCATGCATCTGTCCTCACACTCCTCATCTTTAAGTTATCCTACTGTGGCCCCAATGAAG
TGGACAATTTTTTCTGTGATATTCCGGTGGTGCTGCCCCTGGCCTGCGCAGACACCTCTCTAGCACGGACAGTGAGTT
TCATCAACGTAGGTGTTGTTGCGCTCATGTGTTTTCTTCTTATCCTCACCTCTTATGCTTGCATTGTTATCTCTATAC
TGAAAATCAGTTCCTCAGAAGGTAGGCGCAGAGCCTTCTCAACCTGCAGTGCCCATCTGACGTCCATCCTGCTCTTCT
ATGGACCAATAGTCCTCATTTATCTCCGACCTGCCTCCAGCCCTTGGCTGGACTCTGTGGTTCAGGTGTTGAATAATA
TTGTTATCCCTTCCCTGAATCCTTTGATATATACTTTGAGAAACAAAGGTGTAAAGCTGGCACTGAGAAAGGTGCTCA
TTCAAGGAGTACATAATTGTGGAAGGTAAGCTTTATATCCTGGTCTGTCTTTACAGCTTTGTCTTGAATTTATCTGGG
GGCAATTAGCTCCTATATTTTGTTT (SEQ ID NO:1)

TABLE 3

MKIKNHTPVTEVPLMGIPHTKGMENVLFVLFLAFYLFTLLGNLLILLAVLTFSNLHTPMYFFLGNLSVFDIFFPSVSS
PKMMLCLVGQSCTISFQGCASQLFFHHFLGCTECFLYTVMAYDRFAAICHPLPYTVIMKRRVCALLTLGTWTGSCLHA
SVLTLLIFKLSYCGPNEVDNFFCDIPVVLPLACADTSLARTVSFINVGVVALMCFLLILTSYACIVISILKISSSEGR
RRAFSTCSAHLTSILLFYGPIVLIYLRPASSPWLDSVVQVLNNIVIPSLNPLIYTLRNKGVKLALRKVLIQGVHNCGR
(SEQ ID NO:2)

In a search of sequence databases, it was found, for example, that the SEQ ID NO:1 (Query) has 556 of 890 bases (62%) identical to a Rat species GPCR mRNA (GENBANK-ID: U50949) (Subject) (Table 4). The full amino acid sequence of the protein of the invention(SEQ ID NO:2, Query) was found to have 157 of 305 amino acid residues (51%) identical to, and 202 of 305 residues (66%) positive with, the 318 amino acid residue protein from Rat (ptnr:SPTREMBL-ACC:Q62944) (Subject) (Table 5).

TABLE 4

```
gb:GENBANK-ID:RNU50949|acc:U50949 Rattus norvegicus taste bud receptor protein
TB 641 (TB 641) gene, complete cds - Rattus norvegicus, 1278 bp. Length = 1278
  Minus Strand HSPs:
 Score = 1176 (176.4 bits), Expect = 3.6e-47, P = 3.6e-47
 Identities = 556/890 (62%), Positives = 556/890 (62%), Strand = Minus / Plus

Query:   997 ACACTCCAGTAACTGAGGTCCCCCTGATGGGAATCCCTCATACAAAGGGGATGGAA-AAT 939
             ACACT   GT AC GA  TCC  CT  TGGG  T  CTCAT C       AT  AA AA
Sbjct:   164 ACACTGTGGTGACAGATTTCCTTCTCCTGGGCTTGGCTCATCCCCCAA--ATCTAAGAAC 221

Query:   938 GTGCTTTTTG-TCTTATTTCTGGCCTTCTACCTCTTCACCTTGCTGGGGAACCTACTCAT 880
             GT C T TT  T  T TT CT   C T TAC TC T AC   G TGGGGAACCT CTCAT
Sbjct:   222 GTTCCTCTTCCTGGTCTTCCTCCTCATTTACATCCTGACACAGTTGGGGAACCTGCTCAT 281

Query:   879 TCTTCTGGCCGTCCTCACTTTCTCCAACCTCCA--CAC-CCCCATGTAT-TTCTTCCTGG 824
              CT CT  C GT     CT  C CCAA CT CA  C C CCCCATGTA  TTCT C TGG
Sbjct:   282 CCTGCTCACAGTGTGGGCTGACCCCAAGCTGCATGCCCGCCCCATGTACATTCTGC-TGG 340
```

```
Query:   823 GAAACCTGTCTGTGTTTGACATATTTTTCCCTTCAGTGAGT-TCCCCCAAAATGATGCTC 765
             G    CT TC  T  T GACAT T   TC C TCAGT A T TCCC C AA T AT  T
Sbjct:   341 GCGTGCTCTCCTTCCTGGACATGTGGCTCTCCTCAGTCATTGTCCCTCGAATT-ATTTTA 399

Query:   764 TGCTT-AGTGGGACAAAGCTGCACCATCTCTTTCCAGGGTTGTGCCTCCCAGCTCTTCTT 706
               CTT A T    C AA   G  C ATC C TT    GG TGTG   C CA CTCT  TT
Sbjct:   400 AACTTCACTCCTGCCAACAAGG-CTATCGCATTTGGTGGCTGTGTAGCTCAACTCTATTT 458

Query:   705 TCACCATTTCCTGGGTTGCACCGAGTGTTTCCTGTACACTGTGATGGCCTATGACCGATT 646
             T  CCA TTCCTGGG  GCACC AGTG TTCCT TA AC  TGATGGCCTATGAC G T
Sbjct:   459 TTTCCACTTCCTGGGCAGCACCCAGTGCTTCCTCTATACCTTGATGGCCTATGACAGGTA 518

Query:   645 TGCAGCCATCTGCCACCCTTTGCCATACACGGTCATCATGAAACGCAGGGTGTGTGCCCT 586
                 GC AT TG CA CCT T C  TAC C GT  TCATGAA  G A G T TG  C  T
Sbjct:   519 CCTGGCAATATGTCAGCCTCTTCGCTACCCTGTGCTCATGAATGGGAAGTTATGCACAAT 578

Query:   585 CCTGACG-CTAGGCACCTGGACGGGGAGCTGTCTGCATGCATCTGTCCTCAC-ACTCCTC 528
             CCTG  G CT GG  C TGG  GG   GCT   T CATG  TCT T C   C ACTC T
Sbjct:   579 CCTGGTGGCT-GGAGCTTGGGTGGCTGGCTCCATCCATGGGTCTATTCAAGCCACTC-TG 636

Query:   527 ATCTTTAAGTTATCCTACTGTGGCCCCAATGAAGTGGACAATTTTTTCTGTGATATTCCG 468
             A CTT    TT  CCTACTGTGG CC AA GAAGTGGA  A TT TTCTGTGA ATTCC
Sbjct:   637 ACCTTCCGATTGCCCTACTGTGGGCCTAAGGAAGTGGATTACTTCTTCTGTGACATTCCT 696

Query:   467 GTGGTGCTGCCCCTGGCCTGCGCAGACACCTCTCTAGCACGGACAG-TGAGTTTCATCAA 409
             G  GTGCTG   CTGGCCTG GC GA AC  C  T   A G AC G TGA  TT  T  A
Sbjct:   697 GCAGTGCTGAGACTGGCCTGTGCTGATACAGCAATCA-ATGAACTGGTGACCTTTGTGGA 755

Query:   408 CGTAGGTGTTGTTGC-GCTCATGTGTTTTCTTCTTATCCTCACCTCTTATGCTTGCATTG 350
             C T GG GT GT GC GC CA  TG TT CT CT AT CT   CTC TA GC   CAT G
Sbjct:   756 CATTGGGGTAGTGGCTGC-CAGTTGCTTCCTGCTGATTCTGCTCTCCTACGCCAACATAG 814

Query:   349 TT-ATCTCTATACTGAAAATCAGTTCCTCAGAAGGTAGGCGCAGAGCCTTCTCAACCTGC 291
             TT AT  C AT CTGAA AT  G  C  CAGA GG AGG G  G GCCTTCTC ACCTG
Sbjct:   815 TTCATGCC-ATCCTGAAGATACGCACTGCAGATGGCAGGAGACGTGCCTTCTCCACCTGT 873

Query:   290 AGTGCCCATCTGAC-GTCCATC-CTGCTCTTCTATGGACCAA-TAGTCCTCATTTATCTC 234
              G  CCCATCT AC GT   TC C G TCT CTATG  CC   TA T  TCAT TA CT
Sbjct:   874 GGCTCCCATCTCACTGTGG-TCACAG-TCTACTATGTCCCCTGTATTT-TCATCTACCTT 930

Query:   233 CGACCTGCCTCCAGCCCTTGGCTGGACTCTGTGGTTCAGGTGTTGAATAATATTGTTA-T 175
             CG  C G  TCCA    TT   T GAC   G  GTT   GT TT  A A T TTGT A T
Sbjct:   931 CGGGCAGGTTCCAAGAGTTCCTTTGACGGAGCAGTTGCTGTATTTTACACTGTTGTCACT 990

Query:   174 CCCTTCCCTGAATCCTTTGATATATACTTTGAGAAACAAAGGTGTAAAG-CTGGCACTGA 116
             CC TT C TGAATCC  T AT TA ACT TGAG AAC A G  GT AA  CTG C CTGA
Sbjct:   991 CCATTAC-TGAATCCCCTCATCTACACTCTGAGGAACCAGGAAGTGAATTCTGCC-CTGA 1048

Query:   115 GAAAGG 110 (SEQ ID NO:37)
              A AGG
Sbjct:  1049 -AGAGG 1053 (SEQ ID NO:38)
```

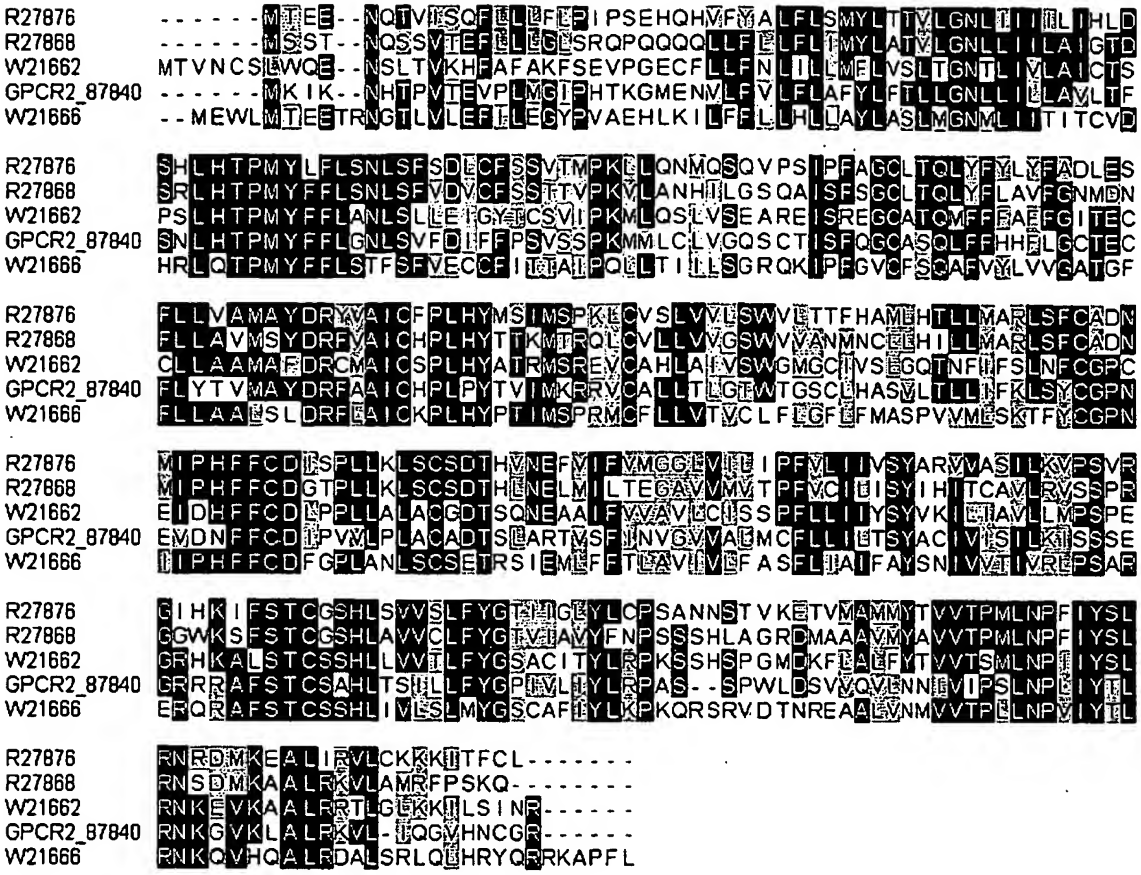
TABLE 5

ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - RATTUS NORVEGICUS (RAT), 318 aa. Length = 318
Minus Strand HSPs:

```
 Score = 799 (281.3 bits), Expect = 9.9e-79, P = 9.9e-79
 Identities = 157/305 (51%), Positives = 202/305 (66%), Frame = -3

Query:   995 TPVTEVPLMGIPHTKGMENVLFVLFLAFYLFTLLGNLLILLAVLTFSNLHT-PMYFFLGN 819
             T VT+  L+G+ H   +   LF++FL  Y+ T LGNLLILL V     LH  PMY  LG
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   818 LSVFDIFFPSVSSPKMMLCLVGQSCTISFQGCASQLFFHHFLGCTECFLYTVMAYDRFAA 639
             LS  D++  SV  P+++L     +  I+F GC +QL+F HFLG T+CFLYT+MAYDR+ A
Sbjct:    71 LSFLDMWLSSVIVPRIILNFTPANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYLA 130

Query:   638 ICHPLPYTVIMKRRVCALLTLGTWTGSCLHASVLTLLIFKLSYCGPNEVDNFFCDIPVVL 459
             IC PL Y V+M  ++C +L  G W     +H S+   L F+L YCGP EVD FFCDIP VL
Sbjct:   131 ICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAVL 190

Query:   458 PLACADTSLARTVSFINVGVVALMCFLLILTSYACIVISILKISSSEGRRRAFSTCSAHL 279
              LACADT++   V+F+++GVVA  CFLLIL SYA IV +ILKI +++GRRRAFSTC +HL
Sbjct:   191 RLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSHL 250

Query:   278 TSILLFYGPIVLIYLRPASSPWLDSVVQVLNNIVIPSLNPLIYTLRNKGVKLALRKVLIQ 99
             T + ++Y P + IYLR  S    D  V V   +V P LNPLIYTLRN+ V  AL++ L
Sbjct:   251 TVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKR-LRA 309

Query:    98 GVHNCG 81 (SEQ ID NO:39)
             G  N G
Sbjct:   310 GRGNVG 315 (SEQ ID NO:40)
```

A multiple sequence alignment is given in Table 6, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. In Table 6, GPCR2_87840 is the full-length SEQ ID NO:2. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers. Based on its relatedness to members of the GPCR family, the NOV1 protein is a novel member of the OR family.

TABLE 6



Cellular localization analysis using Psort of the NOV1 protein of invention indicated that it might be targeted to the plasma membrane (Table 7).

TABLE 7

```
plasma membrane --- Certainty=0.6000(Affirmative) < succ>
Golgi body --- Certainty=0.4000(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3000(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
```

SignalP analysis of NOV1 indicated that it has got secretory signal peptide (Table 8). The first 70 amino acids of NOV1 (312 aa) were used for signal peptide prediction.

TABLE 8

```
< Is the sequence a signal peptide?

# Measure  Position  Value  Cutoff  Conclusion
```

```
max. C    42        0.454    0.37    YES

max. Y    54        0.437    0.34    YES

max. S    36        0.993    0.88    YES

mean S     1-53     0.587    0.48    YES

# Most likely cleavage site between pos. 53 and 54: TFS-NL
```

The results of a search for the homology of SEQ ID NO:1 in the sequence databases using BLASTX is shown in Table 9. These results illustrate a very small P value when comparing the homology of NOV1 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV1 is a novel protein member of the GPCR family.

TABLE 9

Query= SEQ ID NO:1 AC019108_B Cura_145 GPCR (1039 letters)
Translating both strands of query sequence in all 6 reading frames
Database: /opt/database/licensed/blast/geneseq_aa
354,275 sequences; 52,135,959 total letters.

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Probability P(N)	N
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ...	+3	776	2.5e-76	1
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ...	+3	776	2.5e-76	1
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ...	+3	759	1.6e-74	1
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ...	+3	759	1.6e-74	1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ...	+3	733	9.0e-72	1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ...	+3	733	9.0e-72	1
patp:Y92364 G protein-coupled receptor protein 4 - Hom ...	+3	670	4.3e-65	1
patp:R27868 Odorant receptor clone F5 - Rattus rattus, ...	+3	662	3.0e-64	1
patp:W21662 Rat spermatid chemoreceptor D-2 - Rattus s ...	+3	654	2.1e-63	1
patp:R27872 Odorant receptor clone I7 - Rattus rattus, ...	+3	652	3.4e-63	1
patp:W21664 Rat spermatid chemoreceptor D-8 - Rattus s ...	+3	648	9.1e-63	1
patp:W21665 Rat spermatid chemoreceptor D-9 - Rattus s ...	+3	647	1.2e-62	1
patp:R27876 Odorant receptor clone I15 - Rattus rattus ...	+3	644	2.4e-62	1
patp:R27874 Odorant receptor clone I9 - Rattus rattus, ...	+3	637	1.3e-61	1

The highest extent of similarity in the above list is 49%.

NOV2

A novel nucleic acid was identified by TblastN using CuraGen Corporation's sequence file for GPCR run against the Genomic Daily Files made available by GenBank. The nucleic acid was further predicted by the program GenScan™, including selection of exons. These were further modified by

means of similarities using BLAST searches. The sequences were then manually corrected for apparent inconsistencies, thereby obtaining the sequences encoding the full-length protein. Accordingly, a NOV2 sequence according to the invention is a nucleic acid sequence encoding a polypeptide related to the human odorant receptor (OR) family of the G-protein coupled receptor (GPCR) superfamily of proteins.

A NOV2 nucleic acid and its encoded polypeptide includes the sequences shown in Table 10. An open reading frame was identified beginning with an atg initiation codon at nucleotides 47-49 and ending with a taa codon at nucleotides 989-991. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 10, and the start and stop codons are in bold letters. The encoded protein having 314 amino acid residues is presented using the one-letter code in Table 11.

TABLE 10

GAATTTGACTTTTTCTCTCTCATCTCCACAGATTTCTCAGAGAAGA**ATG**GGTGTAAAAAACCATTCCACAGTGACTGA
GTTTCTTCTTTCAGGATTAACTGAACAAGCAGAGCTTCAGCTGCCCCTCTTCTGCCTCTTCTTAGGAATTTACACAGT
TACTGTGGTGGGAAACCTCAGCATGATCTCAATTATTAGGCTGAATCGTCAACTTCATACCCCCATGTACTATTTCCT
GAGTAGTTTGTCTTTTTTAGATTTCTGCTATTCTTCTGTCATTACCCCTAAAATGCTATCAGGGTTTTTATGCAGAGA
TAGATCCATCTCCTATTCTGGATGCATGATTCAGCTGTTTTTTTTCTGTGTTTGTGTTATTTCTGAATGCTACATGCT
GGCAGCCATGGCCTGCGATCGCTACGTGGCCATCTGCAGCCCACTGCTCTACAGGGTCATCATGTCCCCTAGGGTCTG
TTCTCTGCTGGTGGCTGCTGTCTTCTCAGTAGGTTTCACTGATGCTGTGATCCATGGAGGTTGTATACTCAGGTTGTC
TTTCTGTGGATCAAACATCATTAAACATTATTTCTGTGACATTGTCCCTCTTATTAAACTCTCCTGCTCCAGCACTTA
TATTGATGAGCTTTTGATTTTTGTCATTGGTGGATTTAACATGGTGGCCACAAGCCTAACAATCATTATTTCATATGC
TTTTATCCTCACCAGCATCCTGCGCATCCACTCTAAAAAGGGCAGGTGCAAAGCGTTTAGCACCTGTAGCTCCCACCT
GACAGCTGTTCTTATGTTTTATGGGTCTCTGATGTCCATGTATCTCAAACCTGCTTCTAGCAGTTCACTCACCCAGGA
GAAAGTATCCTCAGTATTTTATACCACTGTGATTCTCATGTTGAATCCCTTGATATATAGTCTGAGGAACAATGAAGT
AAGAAATGCTCTGATGAAACTTTTAAGAAGAAAAATATCTTTATCTCCAGGA**TAA**ATATGCTCTTTATTAAGATCTAT
TTCTGTATTCATAATCATGATTATAT SEQ ID NO:3

TABLE 11

MGVKNHSTVTEFLLSGLTEQAELQLPLFCLFLGIYTVTVVGNLSMISIIRLNRQLHTPMYYFLSSLSFLDFCYSSVIT
PKMLSGFLCRDRSISYSGCMIQLFFFCVCVISECYMLAAMACDRYVAICSPLLYRVIMSPRVCSLLVAAVFSVGFTDA
VIHGGCILRLSFCGSNIIKHYFCDIVPLIKLSCSSTYIDELLIFVIGGFNMVATSLTIIISYAFILTSILRIHSKKGR
CKAFSTCSSHLTAVLMFYGSLMSMYLKPASSSSLTQEKVSSVFYTTVILMLNPLIYSLRNNEVRNALMKLLRRKISLS
PG SEQ ID NO:4

In a search of sequence databases, it was found, for example, that the nucleic acid sequence (SEQ ID NO:3, Query) has 697 of 1033 bases (67%) identical to a Rat species GPCR mRNA (GENBANK-ID: X80671) (subject) (Table 12). The full amino acid sequence of the NOV 2 protein of the invention (SEQ ID NO: 4, Query) was found to have 193 of 301 amino acid residues (64%) identical to, and 231 of 301 residues (76%) positive with, the 309 amino acid residue protein from Rat species (ptnr:SPTREMBL-ACC:Q63395, Subject) (Table 13).

TABLE 12

```
gb:GENBANK-ID:RNOLP4|acc:X80671 R.norvegicus olp4 mRNA - Rattus norvegicus,
1441 bp (RNA). Length = 1441
        Plus Strand HSPs:
       Score = 1913 (287.0 bits), Expect = 1.6e-80, P = 1.6e-80
       Identities = 697/1033 (67%), Positives = 697/1033 (67%), Strand = Plus
/ Plus

Query:     8 ACTTT-TTCTCTCT-CATCTCCACAGATTTCTCAGAGAAGAATGGGTGTAAAAAACCATT 65
             ACTTT TT TCT T C    CC  AGA  TCT A A A GA TGGG       AA CATT
Sbjct:   168 ACTTTGTTTTCTGTGCCAGACCT-AGAAATCT-ACAAATGA-TGGGCACTGGGAATCATT 224

Query:    66 CCACAGTGACTGAGTTTCTTCTTTCAGG-ATTAACTGAACAAGCAGAGCTTCAGCTGCCC 124
             C  CAGT   TG GTTT T CTT   GG ATTAAC  A CA  C GAGCT C G TGCCC
Sbjct:   225 CTGCAGTAGTTGTGTTTGT-CTTGGTGGGATTAACAAAGCAGCCTGAGCTCCTGTTGCCC 283

Query:   125 CTCTTCTGCCTCTTCTTAGGAATTTACACAGTT-ACTGTGGTGGGAAACCTCAGCATGAT 183
             CT TTCT CCT TTC T G  AT TA    GTT AC GT GTGGG AA CT  GCATGAT
Sbjct:   284 CTGTTCTTCCTGTTCCTGGTCATCTATGT-GTTGACAGTAGTGGGGAATCTGGGCATGAT 342

Query:   184 CTCAATTATTAGGCTGAATCGTCAACTTCATACCCCCATGTACTATTTCCTGAGTAGTTT 243
             C    T AT A   T A  C  C ACT CA AC CCCATGTACTATTTCCT AG AG TT
Sbjct:   343 CCTGCTCATCATAGTCAGCCCACTACTGCACACTCCCATGTACTATTTCCTCAGCAGCTT 402

Query:   244 GTCTTTTTTAGATTTCTGCTATTCTTCTGTCATTACCCCTAAAATGCTA-TCAGGGTTTT 302
              TC TTT T GAT TCTGCTATTC  CTGTCATTAC CC AAAATGCT  T A   T  T
Sbjct:   403 ATCCTTTGTTGATCTCTGCTATTCCACTGTCATTACACCCAAATGCTGGTGAACTTCCT 462

Query:   303 TATGCAGAGATAGATCCATCTCCTATTCTGGA-TGCATGATTCAGCTGTTTTT-TTT-CT 359
             T  G AGA A A  TC AT  CCTATTC GGA TGCATG   CAG T TTTTT TTT C
Sbjct:   463 TGGGAAGAAAAATTTC-ATTACCTATTC-GGAGTGCATGGCCCAGTTCTTTTTCTTTGCA 520

Query:   360 GTGTTTGTGTTATTTCTGAATGCTACATGCTGGCAGCCATGGCCTGCGATCGCTACGTGG 419
              T TTTGTG T    CTGA  GCTAC T CTG C G  ATGGC T  GATCGCTA GTGG
Sbjct:   521 ATCTTTGTGGTCA--CTGAGGGCTACCTCCTGACTGTTATGGCATATGATCGCTATGTGG 578

Query:   420 CCATCTGCAGCCCACTGCTCTACAGGGTCATCATGTCCCCTAGGGTCTGTTCTCTGCTGG 479
             CCATCTGCAG CCA TGCTCTA A  GT ATCATGTCC CTAG  TCTG TC CTG T G
Sbjct:   579 CCATCTGCAGACCATTGCTCTATAATGTGATCATGTCCTCTAGAATCTGCTCACTGTTAG 638

Query:   480 TGGCTGCT-GTCTTCTCAGTAGGTTTCACTGATGCTGTGATCCATGGAGGTTGTATACTC 538
             TG CTG T G CTTCTC  TAGG  T   T  TGCTGTG T CA   A GT  TAT  T
Sbjct:   639 TG-CTGGTTGCCTTCTCCCTAGGCCTTTTTTCTGCTGTGGTACACACAAGTGCTATGATG 697
```

```
Query:  539 AGGTTGTCTTTCTGTGGATCAAACATCATTAAACATTATTTCTGTGACATTGTCCCTCTT 598
            A   TG   TTCTGT  ATC  ACATCAT A  CATTA TTCTGTGA   T T CC CT
Sbjct:  698 AATCTGAGCTTCTGTAAATCGTACATCATAAGCCATTACTTCTGTGATGCTCTTCCCCTC 757

Query:  599 ATTAAACTCTCCTGCTCCAGCACTTATATTGATGAGCTTTTGATTTTTGTCATTGGTGGA 658
             T AAACTC CCTG TC A CAC  AT T  ATGAGCTT T AT TTT TCATTGG GG
Sbjct:  758 CTCAAACTCGCCTGTTCTAACACACATCTCAATGAGCTTCTCATATTTATCATTGGGGGG 817

Query:  659 TTTAACATGGTGGCCACAAGCCTAACAATCATTATTTCATATGCTTTTATCCTCACCAGC 718
             T AACA   TGG   C A CCTA CA T    AT TC TATG  TT AT  TC  CAGC
Sbjct:  818 CTCAACACCTTGGTGCCCACCCTAGCAGTTGCCATCTCCTATGTCTTCATTTTCTGCAGC 877

Query:  719 ATCCTGCGCATCCACTCTAAAAAGGGCAGGTGCAAAGCGTTTAGCACCTGTAGCTCCCAC 778
            ATCCTGCGCATC   TC   A AGGGCAGGT  AAAGC TTT G ACCTG AGCTC CA
Sbjct:  878 ATCCTGCGCATCAGGTCATCAGAGGGCAGGTCTAAAGCATTTGGAACCTGCAGCTCTCAT 937

Query:  779 CTGACAGCTGTTCTTATGTTTTATGGGTCTCTGATGTCCATGTATCTCAAACCTGCTTCT 838
            CT A  GCTGT    AT TT T TGGGTCT T A  T CATGTAT T AA CCT CTTC
Sbjct:  938 CTCATGGCTGTGGGGATCTTCTTTGGGTCTATCACCTTCATGTATTTAAAGCCTTCTTCA 997

Query:  839 AGCAGTTCACTCACCCAGGAGAAAGTATCCTCAGTATTTTATACCACTGTGATTCTCATG 898
            AG A  TC CT    CA GAGAA GT TC TC GT TT TATACCAC GTGAT C CATG
Sbjct:  998 AGTAACTCTCTGGAGCAAGAGAAGGTGTCTTCTGTGTTCTATACCACAGTGATCCCCATG 1057

Query:  899 TTGAATCCCTTGATATATAGTCTGAGGAACAATGAAGT-AAGAAATGCTCTGATGAAACT 957
             TGAA CC TT ATATATAGT TGAGGAACAA GA GT AAGAAA GC CTG   A A T
Sbjct: 1058 CTGAACCCATTAATATATAGTTTGAGGAACAAAGATGTGAAGAAA-GCACTGGGCAGA-T 1115

Query:  958 TTTAAGA-AGAAAAATA-TCTTTATCTCCAGGATAAATATGCTCTTTATTAAGATCTATT 1015
            T T AG  AG A    A TCTT  T T   G A A A ATGC C  TA  AAG T  ATT
Sbjct: 1116 TCTCAGTCAGGAGTTAAGTCTTACTTTTT-GCAAACA-ATGCACAGTAG-AAG-TGAATT 1171

Query: 1016 TCTGTATTCATAATCATGATTATAT 1040 SEQ ID NO:45
              T  A T A AATC   ATTATAT
Sbjct: 1172 GTTCCAATGAAAATCC--ATTATAT 1194  SEQ ID NO:46
```

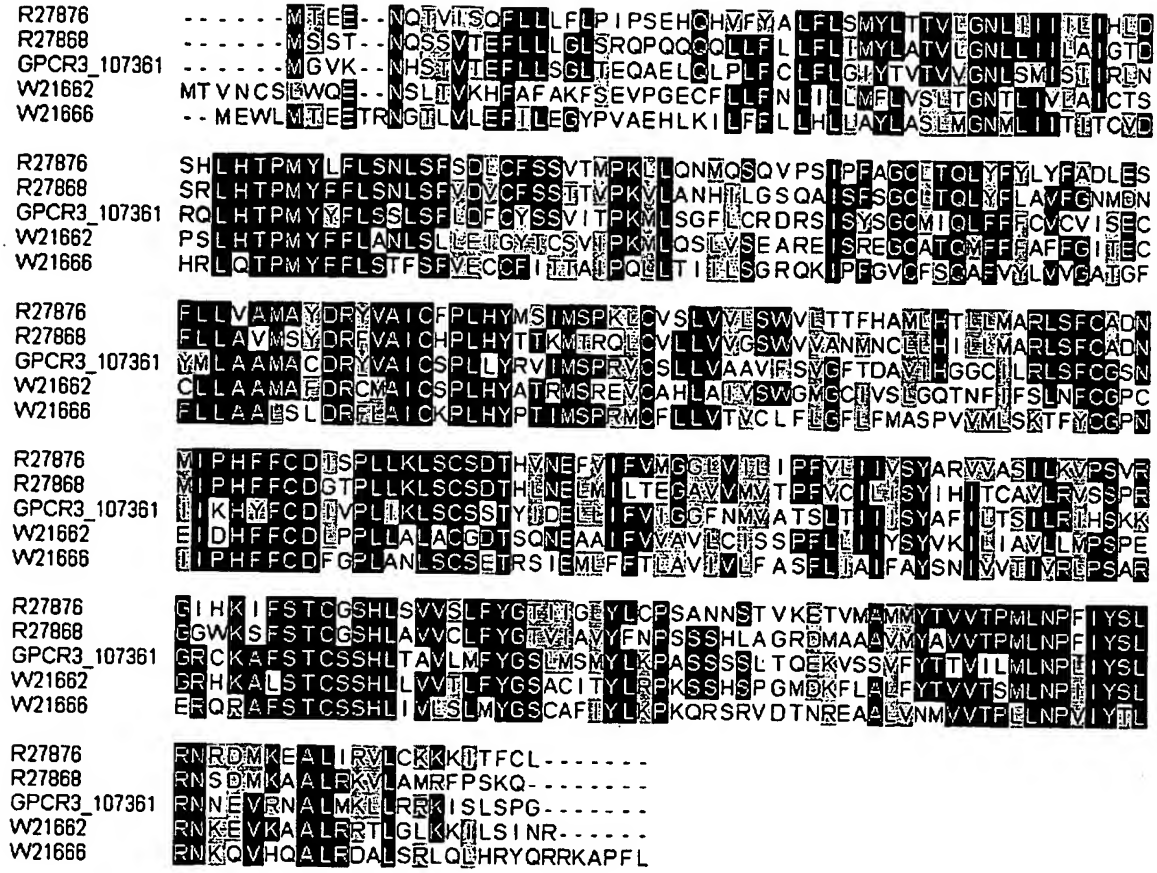
TABLE 13

```
ptnr:SPTREMBL-ACC:Q63395 OLFACTORY RECEPTOR - RATTUS NORVEGICUS (RAT), 309 aa.
Length = 309
Plus Strand HSPs:
 Score = 984 (346.4 bits), Expect = 2.5e-98, P = 2.5e-98
 Identities = 193/301 (64%), Positives = 231/301 (76%), Frame = +2

Query:   1 MGVKNHSTVTEFLLSGLTEQAELQLPLFCLFLGIYTVTVVGNLSMISIIRLNRQLHTPMY 60
           MG  NHS V  F+L GLT+Q EL LPLF LFL IY +TVVGNL MI +I ++  LHTPMY
Sbjct:   2 MGTGNHSAVVVFVLVGLTKQPELLLPLFFLFLVIYVLTVVGNLGMILLIIVSPLLHTPMY 61

Query:  61 YFLSSLSFLDFCYSSVITPKMLSGFLCRDRSISYSGCMIQLFFFCVCVISECYMLAAMAC 120
           YFLSSLSF+D CYS+VITPKML  FL +   I+YS CM Q FFF + V++E Y+L  MA
Sbjct:  62 YFLSSLSFVDLCYSTVITPKMLVNFLGKKNFITYSECMAQFFFFAIFVVTEGYLLTVMAY 121

Query: 121 DRYVAICSPLLYRVIMSPRVCSLLVAAVFSVGFTDAVIHGGCILRLSFCGSNIIKHYFCD 180
           DRYVAIC PLLY VIMS R+CSLLV   FS+G   AV+H   ++ LSFC S II HYFCD
Sbjct: 122 DRYVAICRPLLYNVIMSSRICSLLVLVAFSLGLFSAVVHTSAMMNLSFCKSYIISHYFCD 181
```

```
Query:   181 IVPLIKLSCSSTYIDELLIFVIGGFNMVATSLTIIISYAFILTSILRIHSKKGRCKAFST 240
              +PL+KL+CS+T+++ELLIF+IGG N +  +L + ISY FI  SILRI S +GR KAF T
Sbjct:   182 ALPLLKLACSNTHLNELLIFIIGGLNTLVPTLAVAISYVFIFCSILRIRSSEGRSKAFGT 241

Query:   241 CSSHLTAVLMFYGSLMSMYLKPASSSSLTQEKVSSVFYTTVILMLNPLIYSLRNNEVRNA 300
              CSSHL AV +F+GS+  MYLKP+SS+SL QEKVSSVFYTTVI MLNPLIYSLRN +V+ A
Sbjct:   242 CSSHLMAVGIFFGSITFMYLKPSSSNSLEQEKVSSVFYTTVIPMLNPLIYSLRNKDVKKA 301

Query:   301 L 301  SEQ ID NO:47
              L
Sbjct:   302 L 302  SEQ ID NO:48
```

A multiple sequence alignment is given in Table 14, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. In Table 14, NOV2 (GPCR3_107361) is the full-length SEQ ID NO:4. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers. Based on its demonstrated relatedness to members of the GPCR family, the NOV2 protein is a novel member of the OR family.

TABLE 14



Cellular localization analysis using Psort of the NOV2 protein of invention indicated that it might be targeted to the plasma membrane (Table 15).

TABLE 15

```
plasma membrane --- Certainty=0.6000(Affirmative) < succ>
Golgi body --- Certainty=0.4000(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3000(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
```

SignalP analysis of the NOV2 protein indicated that it has got secretory signal peptide (Table 16). The first 70 amino acids of the NOV2 protein were used for signal peptide prediction.

TABLE 16

```
< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C     42       0.450   0.37   YES
  max. Y     42       0.464   0.34   YES
  max. S     36       0.927   0.88   YES
  mean S      1-41    0.466   0.48   NO
# Most likely cleavage site between pos. 41 and 42: VVG-NL
```

The results of a search for homology of SEQ ID NO:3 in the sequence databases using BLASTX is shown in Table 17. These results illustrate a very small P value when comparing the homology of NOV2 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV2 is a novel protein member of the GPCR family.

TABLE 17

Query= SEQ ID NO:3
(1040 letters)

Translating both strands of query sequence in all 6 reading frames

Database: /opt/database/licensed/blast/geneseq_aa
354,275 sequences; 52,135,959 total letters.

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Probability P(N)	N
patp:Y90875 Human G protein-coupled receptor GTAR11-1 ...	+2	994	2.0e-99	1
patp:Y90875 Human G protein-coupled receptor GTAR11-1 ...	+2	994	2.0e-99	1
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+2	952	5.6e-95	1
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+2	952	5.6e-95	1
patp:Y90876 Human G protein-coupled receptor GTAR11-2 ...	+2	926	3.2e-92	1
patp:Y90876 Human G protein-coupled receptor GTAR11-2 ...	+2	926	3.2e-92	1
patp:Y90878 Human G protein-coupled receptor GTAR11-4 ...	+2	897	3.8e-89	1
patp:Y90878 Human G protein-coupled receptor GTAR11-4 ...	+2	897	3.8e-89	1
patp:Y90879 Human G protein-coupled receptor GTAR11-1 ...	+2	831	3.7e-82	1
patp:Y90879 Human G protein-coupled receptor GTAR11-1 ...	+2	831	3.7e-82	1
patp:Y83394 Olfactory receptor protein OLF-9 - Homo sa...	+2	796	1.9e-78	1
patp:Y83390 Olfactory receptor protein OLF-5 - Homo sa...	+2	791	6.4e-78	1
patp:Y83389 Olfactory receptor protein OLF-4 - Homo sa...	+2	780	9.4e-77	1
patp:Y83387 Olfactory receptor protein OLF-2 - Homo sa...	+2	776	2.5e-76	1

The highest extent of similarity in the above list is 65%.

NOV3

The novel nucleic acid of 1140 nucleotides (designated CuraGen Acc. No. AC019108_D) encoding a novel GPCR-like protein is shown in Table 18. An open reading frame was identified beginning with an ATG intiation codon at nucleotides 36-38 and ending with a TAG codon at nucleotides 969-971. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 1, and the start and stop codons are in bold letters. The encoded protein having 311 amino acid residues is presented using the one-letter code in Table 19.

TABLE 18

TTTCTGTAAGAACAGAGCCCCATATATGAGAAGAAA**ATG**TCCAACGCCACCCTACTGACAGCGTTCATCCTCACGGGCC
TTCCCCATGCCCCAGGGCTGGACGCCCCCCTCTTTGGAATCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACC
TCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCCCCATGTACTACTTCCTCACCAACCTGTCCTTCA
TTGACATGTGGTTCTCCACTGTCACGGTGCCCAAAATGCTGATGACCTTGGTGTCCCCAAGCGGCAGGACTATCTCCT
TCCACAGCTGCGTGGCTCAGCTCTATTTTTTCCACTTCCTGGGGAGCACCGAGTGTTTCCTCTACACAGTCATGTCCT
ATGATCGCTACCTGGCCATCAGTTACCCGCTCAGGTACACCAACATGATGACTGGGCGCTCGTGTGCCCTCCTGGCCA
CCGGCACTTGGCTCAGTGGCTCTCTGCACTCTGCTGTCCAGACCATATTGACTTTCCATTTGCCCTACTGTGGACCCA
ACCAGATCCAGCACTACTTCTGTGACGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCAACGAGATGG
TCATCTTTGTGAATATTGGGCTAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGTGTCCATCGTCTGTT
CCATCCTGCGGATCCGCACCTCAGAGGGGAGGCACAGAGCCTTTCAGACCTGTGCCTCCCACTGTATCGTGGTCCTTT
GCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGCCAGGCTCCAGGGACGCCTTGCATGGGGTTGTGGCCGTTTTCT
ACACCACGCTGACTCCTCTTTTCAACCCTGTTGTGTACACCCTGAGAAACAAGGAGGTAAAGAAAGCTCTGTTGAAGC
TGAAAAATGGGTCAGTATTTGCTCAGGGTGAATAGTTAAGAAAGGCCATATATGGCTAACTTTTCTTTTTTTATTTGT
AATTAAATTAAACCTTCAACATAAGC SEQ ID NO:5

TABLE 19

MSNATLLTAFILTGLPHAPGLDAPLFGIFLVVYVLTVLGNLLILLVIRVDSHLHTPMYYFLTNLSFIDMWFST
VTVPKMLMTLVSPSGRTISFHSCVAQLYFFHFLGSTECFLYTVMSYDRYLAISYPLRYTNMMTGRSCALLA
TGTWLSGSLHSAVQTILTFHLPYCGPNQIQHYFCDAPPILKLACADTSANEMVIFVNIGLVASGCFVLIVLSY
VSIVCSILRIRTSEGRHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRDALHGVVAVFYTTLTPLFNPVVYTLRN
KEVKKALLKLKNGSVFAQGE SEQ ID NO:6

In a search of sequence databases, it was found, for example, that the nucleic acid sequence (SEQ ID NO:5, Query) has 611 of 920 bases (66%) identical to a Rat species GPCR mRNA (GENBANK-ID: U50949) (subject) (Table 20). The full NOV3 amino acid sequence of the protein of

the invention (SEQ ID NO:6, Query) was found to have 170 of 300 amino acid residues (56%) identical to, and 228 of 300 residues (76%) positive with, the 318 amino acid residue protein from Rat (ptnr:SPTREMBL-ACC:Q62944, Subject) (Table 21).

TABLE 20

```
gb:GENBANK-ID:RNU50949|acc:U50949 Rattus norvegicus taste bud receptor protein
TB 641 (TB 641) gene, complete cds - Rattus norvegicus, 1278 bp. Length = 1278
  Plus Strand HSPs:
 Score = 1599 (239.9 bits), Expect.= 2.8e-66, P = 2.8e-66
 Identities = 611/920 (66%), Positives = 611/920 (66%), Strand = Plus / Plus

Query:    32 AGAAGAAATGTCCAACGCCACCCTACTGACAGCGTTCATCCTCACGGGCCTTCCCCATGC 91
             AGAA A A  TC    G CAC  T  TGACAG  TTC T CTC  GGGC T  C CAT C
Sbjct:   148 AGAA-ACACATCACTGGACACTGTGGTGACAGATTTCCTTCTCCTGGGCTTGGCTCATCC 206

Query:    92 CCCAGGGCTG-GA-CGCCCCCCTCTTTGGAATCTTCCTGGTGGTTTACGTGCTCACTGTG 149
             CCCA   CT  GA CG  CC CT   TGG  TCTTCCT  T  TTTAC T CT AC   G
Sbjct:   207 CCCAAATCTAAGAACGTTCCTCTTCCTGG--TCTTCCTCCTCATTTACATCCTGACACAG 264

Query:   150 CTGGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCA--CAC-CCCC 206
             TGGGGAACCT CTCATCCTGCT     T  GGG  GA  C  A CT CA  C C CCCC
Sbjct:   265 TTGGGGAACCTGCTCATCCTGCTCACAGTGTGGGCTGACCCCAAGCTGCATGCCCGCCCC 324

Query:   207 ATGTACTACT-TCCTCACCAACCTGTCCTTCATTGACATGTGGTTCTCCACTGTCACGGT 265
             ATGTAC A T T CT   C   CT TCCTTC T GACATGTGG TCTCC C GTCA  GT
Sbjct:   325 ATGTAC-ATTCTGCTGGGCGTGCTCTCCTTCCTGGACATGTGGCTCTCCTCAGTCATTGT 383

Query:   266 GCCCAAAATGCTGATGACCTTGGTGTCC--CCAAGCGGCAGGACTATCTCCTTCCACAGC 323
              CC   AAT  T  T A CTT    TCC  CCAA C   AGG CTATC C TT     GC
Sbjct:   384 CCCTCGAATTATTTTAAACTTCAC-TCCTGCCAA-CA--AGG-CTATCGCATTTGGTGGC 438

Query:   324 TGCGTGGCTCAGCTCTATTTTTTCCACTTCCTGGGGAGCACCGAGTGTTTCCTCTACACA 383
             TG GT GCTCA CTCTATTTTTTCCACTTCCTGGG AGCACC AGTG TTCCTCTA AC
Sbjct:   439 TGTGTAGCTCAACTCTATTTTTTCCACTTCCTGGGCAGCACCCAGTGCTTCCTCTATACC 498

Query:   384 GTCATGTCCTATGATCGCTACCTGGCCATCAGTTACCCGCTCAGGTACACCAACATGATG 443
              T ATG CCTATGA  G TACCTGGC AT  GT A CC CT  G TAC C     T ATG
Sbjct:   499 TTGATGGCCTATGACAGGTACCTGGCAATATGTCAGCCTCTTCGCTACCCTGTGCTCATG 558

Query:   444 ACTGGGCGCTCGTGTGCCCTCCTGGCCACCGGCACTTGGCTCAGTGGCTCTCTGCACTCT 503
             A TGGG   T  TG  C  TCCTGG   C GG  CTTGG T   TGGCTC  T CA T
Sbjct:   559 AATGGGAAGTTATGCACAATCCTGGTGGCTGGAGCTTGGGTGGCTGGCTCCATCCA-TGG 617

Query:   504 G-CTGTCCAGACCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACCAGATCCAGCA 562
             G CT T CA  CCA   TGAC TTCC  TTGCCCTACTGTGG CC AA  A  T  A  A
Sbjct:   618 GTCTATTCAAGCCACTCTGACCTTCCGATTGCCCTACTGTGGGCCTAAGGAAGTGGATTA 677

Query:   563 CTACTTCTGTGACGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCAACGA 622
             CT CTTCTGTGAC   CC  C  T CTGA ACTGGCCTGTGC GA AC  CA  CAA GA
Sbjct:   678 CTTCTTCTGTGACATTCCTGCAGTGCTGAGACTGGCCTGTGCTGATACAGCAATCAATGA 737

Query:   623 GATGGTCATCTTTGTGAATATTGGGCTAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCT 682
              TGGT A CTTTGTG A ATTGGG TAGTGGC  C  G TGCTT  T CTGAT  TGCT
Sbjct:   738 ACTGGTGACCTTTGTGGACATTGGGGTAGTGGCTGCCAGTTGCTTCCTGCTGATTCTGCT 797
```

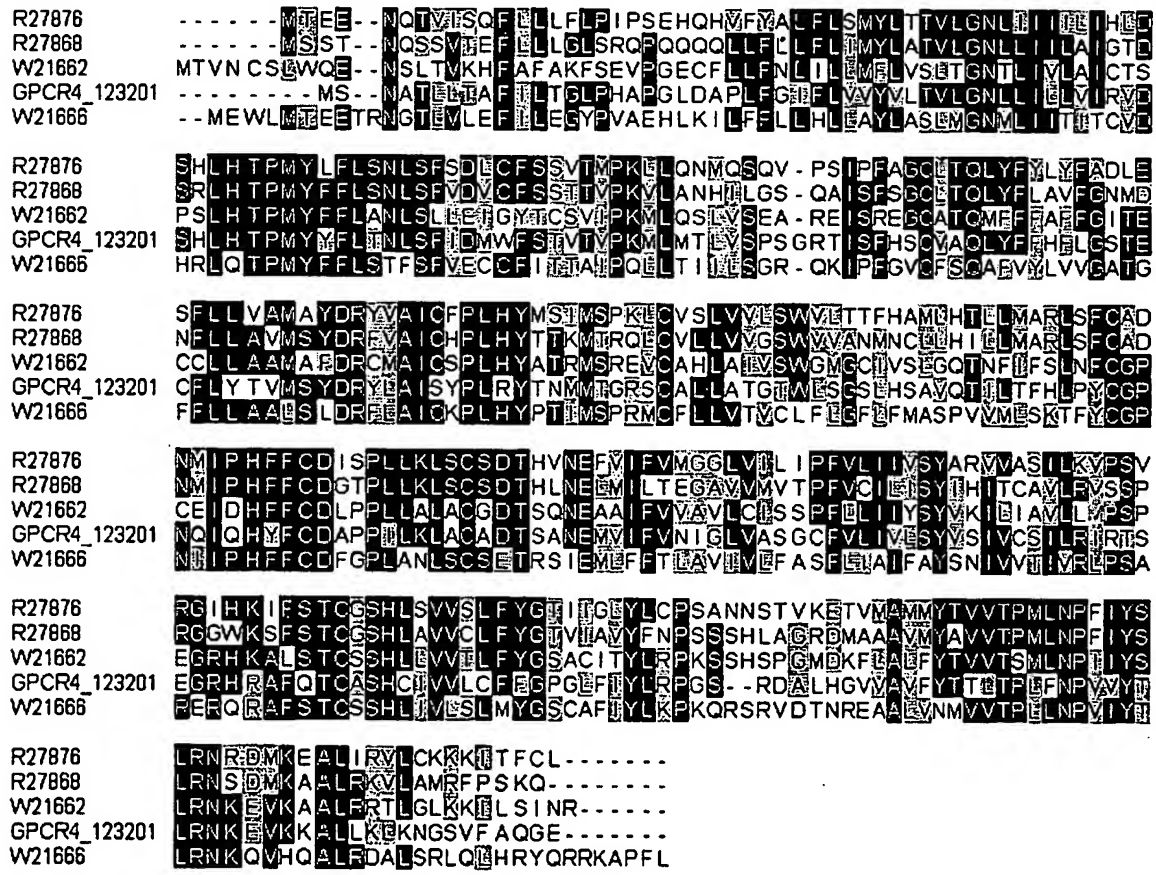
```
Query:   683 GTCCTATGTGTCCATCGTCTGTTCCATCCTGCGGATCCGCACCTCAGAGGGGAGGC-ACA 741
             TCCTA G    CAT GT   T CCATCCTG  GAT CGCAC  CAGA GG AGG  AC
Sbjct:   798 CTCCTACGCCAACATAGTTCATGCCATCCTGAAGATACGCACTGCAGATGGCAGGAGAC- 856

Query:   742 GAGCCTT-TCAGACCTGTGCCTCCCACTGTATC-GTGGTC-CTTTGCTTCTTTGGCCCTG 798
             G GCCTT TC  ACCTGTG CTCCCA T T  C GTGGTC C  T CT CT TG CCC
Sbjct:   857 GTGCCTTCTCC-ACCTGTGGCTCCCA-TCTCACTGTGGTCACAGT-CTACTATGTCCCCT 913

Query:   799 GTCTTTTCATTTACCTGAGGCCAGGCTCCAGGGACGCCTTGCATGGGGTTGTGGCCGTTT 858
             GT TTTTCAT TACCT  GG CAGG TCCA G    CCTT  A GG G  GT GC GT T
Sbjct:   914 GTATTTTCATCTACCTTCGGGCAGGTTCCAAGAGTTCCTTTGACGGAGCAGTTGCTGTAT 973

Query:   859 TCTACACCACGCTGACTCCTCTTTTCAACCCTGTTGTGTACACCCTGAGAAACAAGGAGG 918
             T TACAC     T ACTCC  T  T AA CC  T  T TACAC CTGAG AAC AGGA G
Sbjct:   974 TTTACACTGTTGTCACTCCATTACTGAATCCCCTCATCTACACTCTGAGGAACCAGGAAG 1033

Query:   919 TAAAGAAAGCTCTGTTGAAGCTGAAAAATGGGT 951 (SEQ ID NO:51)
             T AA    GC CTG  GA GCT AA A   GGT
Sbjct:  1034 TGAATTCTGCCCTGAAGAGGCT-AAGAGCAGGT 1065 (SEQ ID NO:52)
```

TABLE 21

```
ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - RATTUS NORVEGICUS
(RAT), 318 aa. Length = 318
Plus Strand HSPs:
Score = 944 (332.3 bits), Expect = 4.3e-94, P = 4.3e-94
Identities = 170/300 (56%), Positives = 228/300 (76%), Frame = +3

Query:    51 TLLTAFILTGLPHAPGLDAPLFGIFLVVYVLTVLGNLLILLVIRVDSHLHT-PMYYFLTN 227
             T++T F+L GL H P L   LF +FL++Y+LT LGNLLILL +  D  LH  PMY  L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   228 LSFIDMWFSTVTVPKMLMTLVSPSGRTISFHSCVAQLYFFHFLGSTECFLYTVMSYDRYL 407
             LSF+DMW S+V VP++++   +P+ + I+F  CVAQLYFFHFLGST+CFLYT+M+YDRYL
Sbjct:    71 LSFLDMWLSSVIVPRIILNF-TPANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYL 129

Query:   408 AISYPLRYTNMMTGRSCALLATGTWLSGSLHSAVQTILTFHLPYCGPNQIQHYFCDAPPI 587
             AI  PLRY  +M G+ C +L  G W++GS+H ++Q  LTF LPYCGP ++ ++FCD P +
Sbjct:   130 AICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAV 189

Query:   588 LKLACADTSANEMVIFVNIGLVASGCFVLIVLSYVSIVCSILRIRTSEGRHRAFQTCASH 767
             L+LACADT+ NE+V FV+IG+VA+ CF+LI+LSY +IV +IL+IRT++GR RAF TC SH
Sbjct:   190 LRLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSH 249

Query:   768 CIVVLCFFGPGLFIYLRPGSRDALHGVVAVFYTTLTPLFNPVVYTLRNKEVKKALLKLKN 947
               VV  ++ P +FIYLR GS+ +  G VAVFYT +TPL NP++YTLRN+EV  AL +L+
Sbjct:   250 LTVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRLRA 309

Query:   948 G 304  SEQ ID NO:53
             G
Sbjct:   310 G 310   SEQ ID NO:54
```

A multiple sequence alignment is given in Table 22, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. These results illustrate a very small P value when comparing the homology of NOV3 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV3 is a novel protein member of the GPCR family.

TABLE 22



Cellular localization analysis using Psort of the NOV3 protein of invention indicated that it might be targeted to the plasma membrane (Table 23).

TABLE 23

```
plasma membrane --- Certainty=0.6400(Affirmative) < succ>
Golgi body --- Certainty=0.4600(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3700(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>
```

SignalP analysis of NOV3 indicated that SEQ ID NO:6 may be a secretory signal peptide (Table 24). The first 70 amino acids of SEQ ID NO:6 (GPCR4_123201 (311 aa)) were used for signal peptide prediction.

TABLE 24

```
< Is the sequence a signal peptide?

# Measure   Position  Value   Cutoff  Conclusion
  max. C     52        0.702   0.37    YES
  max. Y     52        0.650   0.34    YES
  max. S     36        0.982   0.88    YES
  mean S      1-51     0.788   0.48    YES
# Most likely cleavage site between pos. 51 and 52: VDS-HL
```

The results of a search for homology for SEQ ID NO:5 in the sequence databases using BLASTX is shown in Table 25. These results illustrate a very small P value when comparing the homology of NOV3 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV3 is a novel protein member of the GPCR family.

TABLE 25

Query= SEQ ID NO:5
(1040 letters)

Translating both strands of query sequence in all 6 reading frames

Database: /opt/database/licensed/blast/geneseq_aa
354,275 sequences; 52,135,959 total letters.

Sequences producing High-scoring Segment Pairs:	Reading Frame	Smallest Sum High Probability Score	P(N)	N
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ...	+3	923	6.6e-92	1
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ...	+3	923	6.6e-92	1
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ...	+3	884	9.0e-88	1
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ...	+3	884	9.0e-88	1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ...	+3	752	8.7e-74	1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ...	+3	752	8.7e-74	1
patp:R27868 Odorant receptor clone F5 - Rattus rattus, ...	+3	694	1.2e-67	1
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+3	693	1.6e-67	1
patp:Y90877 Human G protein-coupled receptor GTAR11-3 ...	+3	693	1.6e-67	1

The highest extent of similarity in the above list is 56%.

NOV4

The novel nucleic acid of 1040 nucleotides (designated CuraGen Acc. No. AC019108_E, SEQ ID NO:7) encoding a novel GPCR-like protein is shown in Table 26. An open reading frame was identified beginning with an atg initiation codon at nucleotides 61-63 and ending with a tga codon at nucleotides 1030-1032. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 26, and the start and stop codons are in bold letters. The encoded protein having 323 amino acid residues is presented using the one-letter code in Table 27.

TABLE 26

GATTGGTAGCTATAGACCAGTTACAGAGTTCTGTTTCATTTCCCCTGCAGGTGACTGAAG**ATG**AACCCTGAAAACTGG
ACTCAGGTAACAAGCTTTGTCCTTCTGGGTTTCCCCAGTAGCCACCTCATACAGTTCCTGGTGTTCCTGGGGTTAATG
GTGACCTACATTGTAACAGCCACAGGCAAGCTGCTAATTATTGTGCTCAGCTGGATAGACCAACGCCTGCACATACAG
ATGTACTTCTTCCTGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGGTTCCCAAGATGCTTGTCGTC
ATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCATCATCCAGTCCTACCTCTACTTCTTTCTAGGCACCACT
GACTTCTTCCTCTTGGCCGTCATGTCTCTGGATCGTTACCTGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATG
AATGGCCATGTCTGTTCCCAACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCTCTGGGTCCTTTGCCCCACTGTCCTC
ATGGCCAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCTTTCGTGACAGTTGGCCCTTGCTCAGGCTTTCT
TGTGGGGACACCCACCTGCTGAAACTGGTGGCTTTCATGCTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACC
TCAGTTTCCTATGCCTGCATTCTTGCCACTGTTCTCAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGTTTTCCACT

TGCGCCTCGCATCTTACAGTGGTGGTCATCATCTATGGCAGTTCCATCTTTCTCTACATTCGTATGTCAGAGGCTCAG
TCCAAACTGCTCAACAAAGGTGCCTCCGTCCTGAGCTGCATCATCACACCCCTCTTGAACCCATTCATCTTCACTCTC
CGCAATGACAAGGTGCAGCAAGCACTGAGAGAAGCCTTGGGGTGGCCCAGGCTCACTGCTGTGATGAAACTGAGGGTC
ACAAGTCAAAGGAAATGATCTTATTA SEQ ID NO:7

TABLE 27

MNPENWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFLRNFSFLELLLVTVVV
PKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRYLAICRPLRYETLMNGHVCSQLVLASWLAGFLWV
LCPTVLMASLPFCGPNGIDHFFRDSWPLLRLSCGDTHLLKLVAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAER
RKAFSTCASHLTVVVIIYGSSIFLYIRMSEAQSKLLNKGASVLSCIITPLLNPFIFTLRNDKVQQALREALGWPRLTA
VMKLRVTSQRK SEQ ID NO:8

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention (SEQ ID NO:7, Query) has 552 of 900 bases (61%) identical to a gb:GENBANK-ID:AF102523|acc:AF102523.1 mRNA from Mus musculus (Mus musculus olfactory receptor C6 gene, complete cds) (subject) (Table 28). The full amino acid sequence of the protein of the invention (SEQ ID NO:8, Query) was found to have 151 of 300 amino acid residues (50%) identical to, and 213 of 300 amino acid residues (71%) similar to, the 313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus musculus (Mouse) (OLFACTORY RECEPTOR C6) (subject) (Table 29).

TABLE 28

```
gb:GENBANK-ID:AF102523|acc:AF102523 Mus musculus olfactory receptor C6 gene,
complete cds - Mus musculus, 942 bp.  Length = 942
  Plus Strand HSPs:
 Score = 1178 (176.7 bits), Expect = 3.9e-47, P = 3.9e-47
 Identities = 556/900 (61%), Positives = 556/900 (61%), Strand = Plus / Plus

Query:    73 AACTGGACTCAGGTAACA-AGCTTTGTCCTTCTGGGTTTCCCCAGTAGCCACCTCA--TA 129
             AAC G ACT   GT AC  AG TTT T  T CTGGG  T  C AG  GCC   T A  T
Sbjct:     7 AACAGCACTACTGTTACTGAG-TTTATTTTGCTGGGGCTGTC-AGATGCCTG-TGAGCTG 63

Query:   130 CAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAACAGCCACAGGCAAGCTG 189
             CAG T CT  T TTCCTGGG TT  T  TGACCTAC T  T A   C  C GG AA CT
Sbjct:    64 CAGGTGCTCATATTCCTGGGCTTTCTCCTGACCTACTTCCTCATT-CTGCTGGGAAACTT 122

Query:   190 C-TAATTATTGTGCTCAGC-TGGATAGACCAACGCCTGCACATACAGATGTACTTCTTCC 247
             C T AT AT  T  TCA C T G T GAC   CGCCT  ACA  C  ATGTA T CTTCC
Sbjct:   123 CCTCATCATCTTCATCACCCTTG-TGGACAGGCGCCTTTACACCCCCATGTATTACTTCC 181

Query:   248 TGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGGTTCCCAAGATGCTTG 307
             T CG AA TT  CC T CTGGAG T T G T     CTGT  T  T CCCAAGATGCT
```

```
Sbjct:  182 TCCGCAACTTTGCCATGCTGGAGATCTGGTTCACCTCTGTCATCTTCCCCAAGATGCTAA 241

Query:  308 TCGTCATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCATCATCCAGTCCTACC 367
            C  CATC TCAC GG  AT A ACCATCTC  T  T  G TG  TC TCCA  C T CC
Sbjct:  242 CCAACATCATCACAGGACATAAGACCATCTCCCTACTAGGTTGTTTCCTCCAAGCATTCC 301

Query:  368 TCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCTTGGCCGTCATGTCTCTGGATCGTT 427
            TCTA TTCTT CT GGCACCACTGA TTCTT CT  TGGC GT ATGTC  T GA  G T
Sbjct:  302 TCTATTTCTTCCTTGGCACCACTGAGTTCTTTCTACTGGCAGTGATGTCCTTTGACAGGT 361

Query:  428 ACCTGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATGGCCATGTCTGTTCCC 487
            A  TGGC AT TG    CC  T CG TATG  ACC T ATGA        GTCTGT  CC
Sbjct:  362 ATGTGGCCATTTGTAACCCTTTGCGTTATGCCACCATTATGAGCAAAAGAGTCTGTGTCC 421

Query:  488 AACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCT-CTGGGTCCTTTGCCCCACTGTCC-T 545
            A CT GTG T   CTC TGG T  CTGGATT CT CT   TC T    CC A T TC  T
Sbjct:  422 AGCTTGTGTTTTGCTCATGGATGTCTGGATTGCTTCTCA-TCATAGTTCCTAGT-TCAAT 479

Query:  546 CATGGC-CAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCTTTCGTGACAGTT 604
              T    CAGC  GCC TTCTGTGGCCC AA    ATT A CA TTCTT  GTGACA  T
Sbjct:  480 TGTATTTCAGCA-GCCATTCTGTGGCCCAAACATCATTAATCATTTCTTCTGTGACAACT 538

Query:  605 GGCC-CTTGCTCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGGTGGCTTTCATG 663
              CC CTT    A  CT    TGTG  GA AC   CCTG T  A  T  TGG TTT  T
Sbjct:  539 TTCCACTTATGGAA-CTCATATGTGCAGATACTAGCCTGGTAGAGTTCCTGGGTTTTGTT 597

Query:  664 CTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACCTCAGTTTCCTATGCCTGCATT 723
            T  C A  TT     T CTGGGC C CTGGCT TGAC  C    T CTATG C  CATT
Sbjct:  598 ATTGCCAATTTCAGCCTCCTGGGCACTCTGGCTGTGACTGCCACCTGCTATGGCCACATT 657

Query:  724 CTTGCCACTGTTCT-CAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGTTTTCCACTTG 782
            CT    AC  TTCT CA    CC T CAGC    GAG G A GAAAGC TT TC ACTTG
Sbjct:  658 CTCTATACCATTCTACACATTCCTT-CAGCCAAGGAGAGGAAGAAAGCCTTCTCAACTTG 716

Query:  783 CGCCTCGCATCTTACAGTGGTGG-TCATCATCTATGGCAGTTCCATCTTTCTCTACATTC 841
            C CCTC CAT TTA  GTGGTG  TC TC TCTA GGCAG T  ATCTT  T TA  T C
Sbjct:  717 CTCCTCTCATATTATTGTGGTGTCTC-TCTTCTACGGCAGCTGTATCTTCATGTATGTCC 775

Query:  842 G-TATGTCA-GAG-GCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTC-CTGAGCTGC 897
            G T TG CA GA  G  CAG    A    C  AACAA GGTG    C T  CT A C
Sbjct:  776 GGTCTGGCAAGAATGGACAGGGGGAGGATCATAACAA-GGTGGTGGCATTGCTCAACACT 834

Query:  898 ATCATCACACCC-CTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGCAGCA 956
              T  T ACACCC C CT  AACCC TTCATCT CACTCT  G AA  A  AGGTG AGCA
Sbjct:  835 GTAGTGACACCCACACTC-AACCCCTTCATCTACACTCTGAGGAACAAGCAGGTGAAGCA 893

Query:  957 AGCACTGAGAGAA 969   SEQ ID NO:59
              G A T AG GAA
Sbjct:  894 GGTATTTAGGGAA 906   SEQ ID NO:60
```

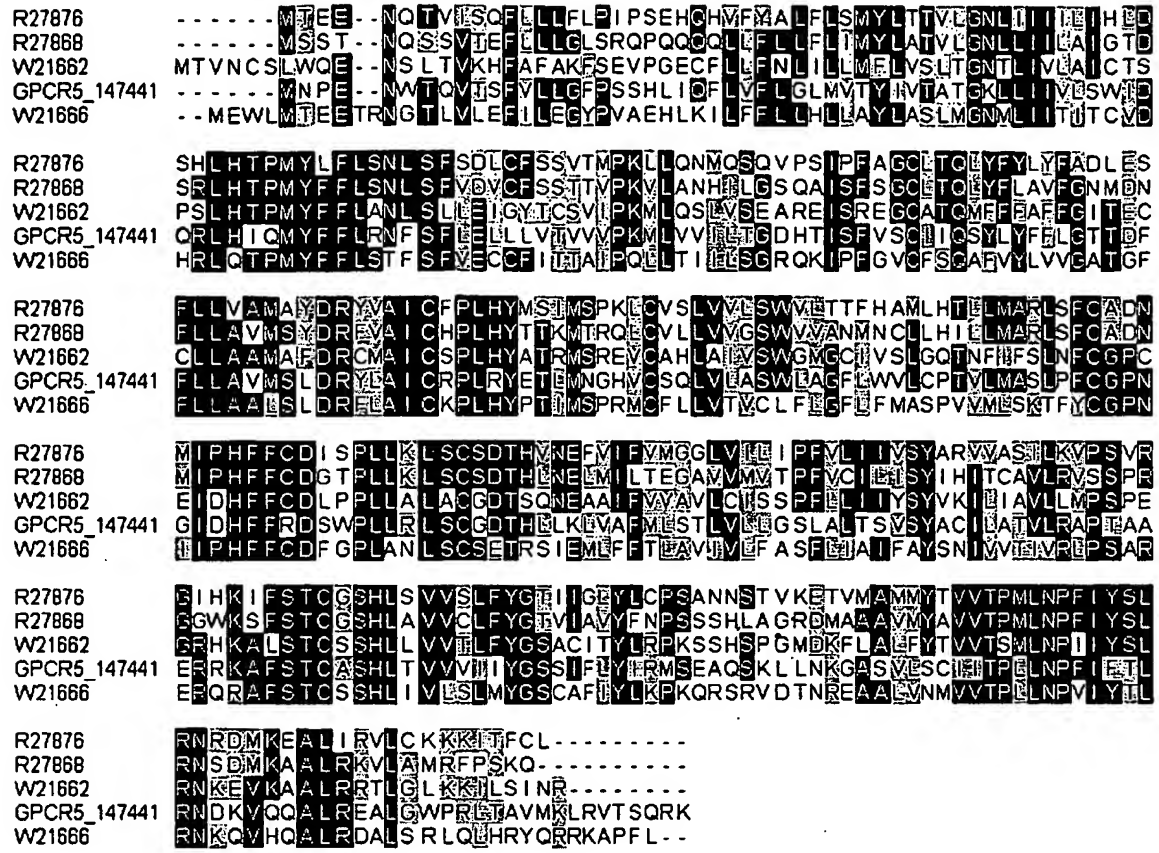
TABLE 29

ptnr:SPTREMBL-ACC:O70271 OLFACTORY RECEPTOR-LIKE PROTEIN - RATTUS NORVEGICUS (RAT), 327 aa. Length = 327
Plus Strand HSPs:

```
Score = 710 (249.9 bits), Expect = 2.7e-69, P = 2.7e-69
Identities = 133/302 (44%), Positives = 204/302 (67%), Frame = +1

Query:    70 ENWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFL 249
             +N T V  F+L GFP +  ++ L FL  M+ Y+ +  G +LII  + +D RL   MYFFL
Sbjct:    10 KNGTLVQEFILEGFPVAEHLRILFFLLHMLAYLASLMGNMLIITYTCVDHRLQTPMYFFL 69

Query:   250 RNFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRY 429
               FSF+E   +T V+P++L +IL+G   I F++C  Q+++   FLG   FFL+AV+SLDR+
Sbjct:    70 STFSFVECCFITTVIPQLLTIILSGRQKIPFMACFSQAFVVLFLGAAVFFLMAVLSLDRF 129

Query:   430 LAICRPLRYETLMNGHVCSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRDSWP 609
             LAIC+PL Y T+M+  +C  LV  S + GFL++  P V+++   +CGPN I HFF D  P
Sbjct:   130 LAICKPLHYPTIMSPRMCFLLVTVSLVLGFLFMASPVVMLSQSFYCGPNIIPHFFCDFGP 189

Query:   610 LLRLSCGDTHLLKLVAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFSTCAS 789
             L  LSC +T  ++++ F L+ +VL  SL +   +Y+ I+ T++R P+A ER++AFSTC+S
Sbjct:   190 LANLSCSETRSIEMLFFTLAIIVLFTSLLIAIFAYSTIVVTIVRLPSARERQRAFSTCSS 249

Query:   790 HLTVVVIIYGSSIFLYIRMSEAQSKLLNKGASVLSCIITPLLNPFIFTLRNDKVQQALRE 969
             HL V+ ++YGS +F+Y++  +       N+ A +++ ++TPLLNP I+TLRN +V QALR+
Sbjct:   250 HLIVLSLMYGSCVFIYLKPKQRSRVDTNREAVLVNMVVTPLLNPVIYTLRNKQVHQALRD 309

Query:   970 AL 975  SEQ ID NO:61
             AL
Sbjct:   310 AL 311  SEQ ID NO:62
```

A multiple sequence alignment is given in Table 30, with the NOV4 protein of the invention being shown on line 4, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV4 protein is a novel member of the OR family. In the following table, GPCR5_147441 is the full-length SEQ ID NO:8. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 30



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name 7tm_1 (7 transmembrane receptor (rhodopsin family)) at amino acid positions 41 to 290. This indicates that the NOV4 sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Tissue expression

NOV4 is expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus

callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs. "Depth" rerepresents the number of clones covering the region of the SNP. The Putative Allele Frequency (Putative Allele Freq.) is the fraction of all the clones containing the SNP. The sign ">" means "is changed to".

Cons.Pos.: 65 Depth: 26 Change: T > C
Putative Allele Freq.: 0.192
Cons.Pos.: 76 Depth: 26 Change: T > C
Putative Allele Freq.: 0.077
Cons.Pos.: 106 Depth: 26 Change: T > G
Putative Allele Freq.: 0.115
Cons.Pos.: 199 Depth: 26 Change: C > T
Putative Allele Freq.: 0.077
Cons.Pos.: 294 Depth: 25 Change: C > T
Putative Allele Freq.: 0.080
Cons.Pos.: 431 Depth: 22 Change: C > G
Putative Allele Freq.: 0.364
Cons.Pos.: 615 Depth: 11 Change: C > G
Putative Allele Freq.: 0.273
Cons.Pos.: 732 Depth: 17 Change: C > T
Putative Allele Freq.: 0.353
Cons.Pos.: 760 Depth: 18 Change: A > G
Putative Allele Freq.: 0.111
Cons.Pos.: 859 Depth: 18 Change: C > T
Putative Allele Freq.: 0.222
Cons.Pos.: 956 Depth: 18 Change: T > A
Putative Allele Freq.: 0.222

Cellular localization analysis using Psort of the protein of invention indicated that NOV4 might be targeted to the plasma membrane (Table 31). The results predict that this sequence has a signal peptide and is likely to be localized on the plasma membrane with a certainty of 0.6400. The first 41 amino acids are more likely to be cleaved as a signal peptide based on the SignalP result.

TABLE 31

```
plasma membrane --- Certainty=0.6400(Affirmative) < succ>
Golgi body --- Certainty=0.4600(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3700(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>
```

SignalP analysis of the NOV4 protein of invention indicated that it has got secretory signal peptide (Table 32). The first 70 amino acids of SEQ ID NO:8 (GPCR5_147441 (323 aa)) were used for signal peptide prediction

TABLE 32

```
< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    39        0.534   0.37   YES
  max. Y    42        0.455   0.34   YES
  max. S    33        0.965   0.88   YES
  mean S     1-41     0.756   0.48   YES
# Most likely cleavage site between pos. 41 and 42: ATG-KL
```

The results of a search for homology for SEQ ID NO:7 in the sequence databases using BLASTX is shown in Table 33.

These results illustrate a very small P value when comparing the homology of NOV4 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV4 is a novel protein member of the GPCR family.

TABLE 33

QUERY= SEQ ID NO:7
(1040 LETTERS)

TRANSLATING BOTH STRANDS OF QUERY SEQUENCE IN ALL 6 READING FRAMES

DATABASE: /OPT/DATABASE/LICENSED/BLAST/GENESEQ_AA
354,275 SEQUENCES; 52,135,959 TOTAL LETTERS.

SMALLEST

SEQUENCES PRODUCING HIGH-SCORING SEGMENT PAIRS:	SUM READING FRAME	HIGH SCORE	PROBABILITY P(N)	N
PATP:Y96680 MURINE OLFACTORY RECEPTOR LIGAND-BINDING R...	+1	826	1.3E-81	1
PATP:W21666 RAT SPERMATID CHEMORECEPTOR G-X - RATTUS S...	+1	689	4.1E-67	1
PATP:R27869 ODORANT RECEPTOR CLONE F6 - RATTUS RATTUS,...	+1	687	6.7E-67	1
PATP:R27872 ODORANT RECEPTOR CLONE I7 - RATTUS RATTUS,...	+1	662	3.0E-64	1

THE HIGHEST EXTENT OF SIMILARITY IN THE ABOVE LIST IS 50%.

NOV5

The novel nucleic acid of 1120 nucleotides (designated CuraGen Acc. No. AC019108_F, SEQ ID NO:9 encoding a novel GPCR-like protein is shown in Table 34. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 60-62 and ending with a TGA codon at nucleotides 1056-1058. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 34, and the start and stop codons are in bold letters. The encoded protein having 332 amino acid residues (SEQ ID NO:10) is presented using the one-letter code in Table 35.

TABLE 34

AAGTATCCTTTTCTGTCCTTATTTGGGAGTGACACAGAATTGATGTGCTGAGCCTAGAAATGTTAGAGGGTGTTGAGC
ATCTCCTTCTGCTACTTCTTTTGACAGATGTGAACAGCAAGGAACTGCAAAGTGGAAACCAGACTTCTGTGTCTCACT
TCATTTTGGTGGGCCTGCACCACCCACCACAGCTGGGAGCGCCACTCTTCTTAGCTTTCCTTGTCATCTATCTCCTCA
CTGTTTCTGGAAATGGGCTCATCATCCTCACTGTCTTAGTGGACATCCGGCTCCATCGTCCCATGTGCTTGTTCCTGT
GTCACCTCTCCTTCTTGGACATGACCATTTCTTGTGCTATTGTCCCCAAGATGCTGGCTGGCTTTCTCTTGGGTAGTA
GGATTATCTCCTTTGGGGGCTGTGTAATCCAACTATTTTCTTTCCATTTCCTGGGCTGTACTGAGTGCTTCCTTTACA
CACTCATGGCTTATGACCGTTTCCTTGCCATTTGTAAGCCCTTACACTATGCTACCATCATGACCCACAGAGTCTGTA
ACTCCCTGGCTTTAGGCACCTGGCTGGGAGGGACTATCCATTCACTTTTCCAAACAAGTTTTGTATTCCGGCTGCCCT
TCTGTGGCCCCAATCGGGTCGACTACATCTTCTGTGACATTCCTGCCATGCTGCGTCTAGCCTGCGCCGATACGGCCA
TCAACGAGCTGGTCACCTTTGCAGACATTGGCTTCCTGGCCCTCACCTGCTTCATGCTCATCCTCACTTCCTATGGCT
ATATTGTAGCTGCCATCCTGCGAATTCCGTCAGCAGATGGGCGCCGCAATGCCTTCTCCACTTGTGCTGCCCACCTCA
CTGTTGTCATTGTTTACTATGTGCCCTGCACCTTCATTTACCTGCGGCCTTGTTCACAGGAGCCCCTGGATGGGGTGG
TAGCTGTCTTTTACACTGTCATCACTCCCTTGCTTAACTCCATCATCTACACACTGTGCAACAAAGAAATGAAGGCAG
CATTACAGAGGCTAGGGGGCCACAAGGAAGTGCAGCCTCACTGACTTCATCCATCACAGACATGCAGGAAACCCTATA
ATGCAGGGCCCTGGATAGGACACCAAAT SEQ ID NO:9

TABLE 35

MLEGVEHLLLLLLLLTDVNSKELQSGNQTSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTVLVDIRLHR
PMCLFLCHLSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMAYDRFLAICKPLHYATI
MTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCDIPAMLRLACADTAINELVTFADIGFLALTCFML

ILTSYGYIVAAILRIPSADGRRNAFSTCAAHLTVVIVYYVPCTFIYLRPCSQEPLDGVVAVFYTVITPLLNSIIYTLC
NKEMKAALQRLGGHKEVQPH SEQ ID NO:10

In a search of sequence databases, it was found, for example, that the nucleic acid sequence(SEQ ID NO:9, Query) has 659 of 976 bases (67%) identical to a Rat species GPCR mRNA (GENBANK-ID: U50949, Subject) (Table 36). The full amino acid sequence of the NOV5 protein of the invention (SEQ ID NO:10, Query) was found to have 185 of 297 amino acid residues (62%) identical to, and 225 of 297 residues (75%) positive with, the 318 amino acid residue protein from Rat (ptnr:SPTREMBL-ACC:Q62944, Subject) (Table 37).

TABLE 36

```
gb:GENBANK-ID:RNU50949|acc:U50949 Rattus norvegicus taste bud receptor protein
TB 641 (TB 641) gene, complete cds - Rattus norvegicus, 1278 bp. Length = 1278
  Minus Strand HSPs:
 Score = 1765 (264.8 bits), Expect = 8.9e-74, P = 8.9e-74
 Identities = 659/976 (67%), Positives = 659/976 (67%), Strand = Minus / Plus

Query:  1041 TCTC-CTTC-TGCTACTTCTTTTGACAGATGTGA-ACA-GC-AAGGAACTGCAAAGTGGA 987
             TCT  CTTC TG T C TCTT   ACAGAT T A ACA G  AAG AAC G AAA
Sbjct:   102 TCTGACTTCATGGTTCATCTTA--ACAGAT-TAAGACATGAGAAGAAACAG-AAACACAT 157

Query:   986 AACCAGACT-TCTG-TGTCTCACTTCATTTTGGTGGGCCTGCACCACCCACCACAGCTGG 929
              AC  GAC  T TG TG C  A TTC TT T  TGGGC TG   CA CC CCA A CT
Sbjct:   158 CACTGGACACTGTGGTGACAGATTTCCTTCTCCTGGGCTTGGCTCATCCCCCAAATCTAA 217

Query:   928 GAGCGCCACTCTTCTTAGCTTTCCTTGTCATCTATCTCCTCACTGTTTCTGGAAATGGGC 869
             GA CG   CTCTTC T G  TTCCT  TCAT TA  TCCT AC    T  GG AA   GC
Sbjct:   218 GAACGTTCCTCTTCCTGGTCTTCCTCCTCATTTACATCCTGACACAGTTGGGGAACCTGC 277

Query:   868 TCATCATCCTCACTGTCTTAGTGGACATCCGGCTCCAT-C--GTCCCATGTGCTTGTTCC 812
             TCATC T CTCAC GT T  G  GAC  C  GCT CAT C  G CCCATGT C T  T C
Sbjct:   278 TCATCCTGCTCACAGTGTGGGCTGACCCCAAGCTGCATGCCCGCCCCATGTACATTCTGC 337

Query:   811 TGTGTCACCTCTCCTTCTTGGACATGACCATTTCTTGTGCTATTGTCCC-CAAGATGCTG 753
             TG G    CTCTCCTTC TGGACATG    T TC T  G  ATTGTCCC C A AT  T
Sbjct:   338 TGGGCGTGCTCTCCTTCCTGGACATGTGGCTCTCCTCAGTCATTGTCCCTCGA-ATTATT 396

Query:   752 GCTGGCTTT-CTCTTGGGTAGTAGGATTATCTCCTTTGGGGGCTGTGTAATCCAACTATT 694
                  CTT  CTC TG   A  AGG T ATC C TTTGG GGCTGTGTA   CAACT T
Sbjct:   397 TTAAACTTCACTCCTGCCAACAAGGCT-ATCGCATTTGGTGGCTGTGTAGCTCAACTCTA 455

Query:   693 TTCTTTCCATTTCCTGGGCTGTACTGAGTGCTTCCTTTACACACTCATGGCTTATGACCG 634
             TT TTTCCA TTCCTGGGC G AC  AGTGCTTCCT TA AC  T ATGGC TATGAC G
Sbjct:   456 TTTTTTCCACTTCCTGGGCAGCACCCAGTGCTTCCTCTATACCTTGATGGCCTATGACAG 515

Query:   633 TTTCCTTGCCATTTGTAAGCC-CTTACACTATGCTACCATCATGACCCACAGAGTC-TGT 576
              T CCT GC AT TGT AGCC CTT C CTA  CT    TCATGA     A AGT  TG
Sbjct:   516 GTACCTGGCAATATGTCAGCCTCTT-CGCTACCCTGTGCTCATGAATGGGA-AGTTATGC 573

Query:   575 A-ACTCCCTGGCTTTAGGCACCTGGCTGGGAGGGACTATCCATTCA-CTTTTCCAAACAA 518
             A A TCC TGG     GG  C TGG TGG  GG  C ATCCAT    CT TTC AA C A
```

```
Sbjct:   574 ACAATCC-TGGTGGCTGGAGCTTGGGTGGCTGGCTCCATCCATGGGTCTATTC-AAGCCA 631

Query:   517 GTTTTGTATTCCGGCTGCCCTTCTGTGGCCCCAATCGGGTCGACTACATCTTCTGTGACA 458
             T T    TTCCG  TGCCCT CTGTGG CC AA    GT GA TAC TCTTCTGTGACA
Sbjct:   632 CTCTGACCTTCCGATTGCCCTACTGTGGGCCTAAGGAAGTGGATTACTTCTTCTGTGACA 691

Query:   457 TTCCTGCCATGCTGCGTCTAGCCTGCGCCGATACGGCCATCAACGAGCTGGTCACCTTTG 398
             TTCCTGC  TGCTG G CT GCCTG GC GATAC GC ATCAA GA CTGGT ACCTTTG
Sbjct:   692 TTCCTGCAGTGCTGAGACTGGCCTGTGCTGATACAGCAATCAATGAACTGGTGACCTTTG 751

Query:   397 CAGACATTGGCTTCCTGGCCCTCACCTGCTTCATGCTCATCCTCACT-TCCTATGGCTAT 339
               GACATTGG  T  TGGC   CA  TGCTTC TGCT AT CT  CT TCCTA G C A
Sbjct:   752 TGGACATTGGGGTAGTGGCTGCCAGTTGCTTCCTGCTGATTCTG-CTCTCCTACGCCAAC 810

Query:   338 ATTGTAGCTGCCATCCTGCGAATTCCGTCAGCAGATGGGCGCCGCAATGCCTTCTCCACT 279
             AT GT   TGCCATCCTG   AT C   C GCAGATGG  G  G   TGCCTTCTCCAC
Sbjct:   811 ATAGTTCATGCCATCCTGAAGATACGCACTGCAGATGGCAGGAGACGTGCCTTCTCCACC 870

Query:   278 TGTG-CTGCCCACCTCACTGTTGTCATTGTTTACTATGTGCCCTGCACCTTCATTTACCT 220
             TGTG CT CCCA CTCACTGT GTCA  GT TACTATGT CCCTG A  TTCAT TACCT
Sbjct:   871 TGTGGCT-CCCATCTCACTGTGGTCACAGTCTACTATGTCCCCTGTATTTTCATCTACCT 929

Query:   219 GCGGCCTTGTTCACAGGAGCCCCTG-GATGGGGTGGTAGCTGTCTTTTACACTGTCATCA 161
              CGG C  GTTC CA GAG  CCT  GA GG G  GT GCTGT TTTTACACTGT  TCA
Sbjct:   930 TCGGGCAGGTTC-CAAGAGTTCCTTTGACGGAGCAGTTGCTGTATTTTACACTGTTGTCA 988

Query:   160 CTCCCTTGCTTAACTCCATCATCTACACACTGTGCAACAAAGAAATGAAGGCAGCATTAC 101
             CTCC TT CT AA  CC TCATCTACAC CTG G AAC A GAA TGAA  C GC  T
Sbjct:   989 CTCCATTACTGAATCCCCTCATCTACACTCTGAGGAACCAGGAAGTGAATTCTGCCCTGA 1048

Query:   100 AGAGGCTAGGGG--GCCACAAGGAA-GTG 75   SEQ ID NO:67
             AGAGGCTA G G  G  A A GGAA GTG
Sbjct:  1049 AGAGGCTAAGAGCAGGTAGAGGGAATGTG 1077  SEQ ID NO:68
```

TABLE 37

ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - RATTUS NORVEGICUS RAT), 318 aa. Length = 318

Minus Strand HSPs:

Score = 969 (341.1 bits), Expect = 9.6e-97, P = 9.6e-97

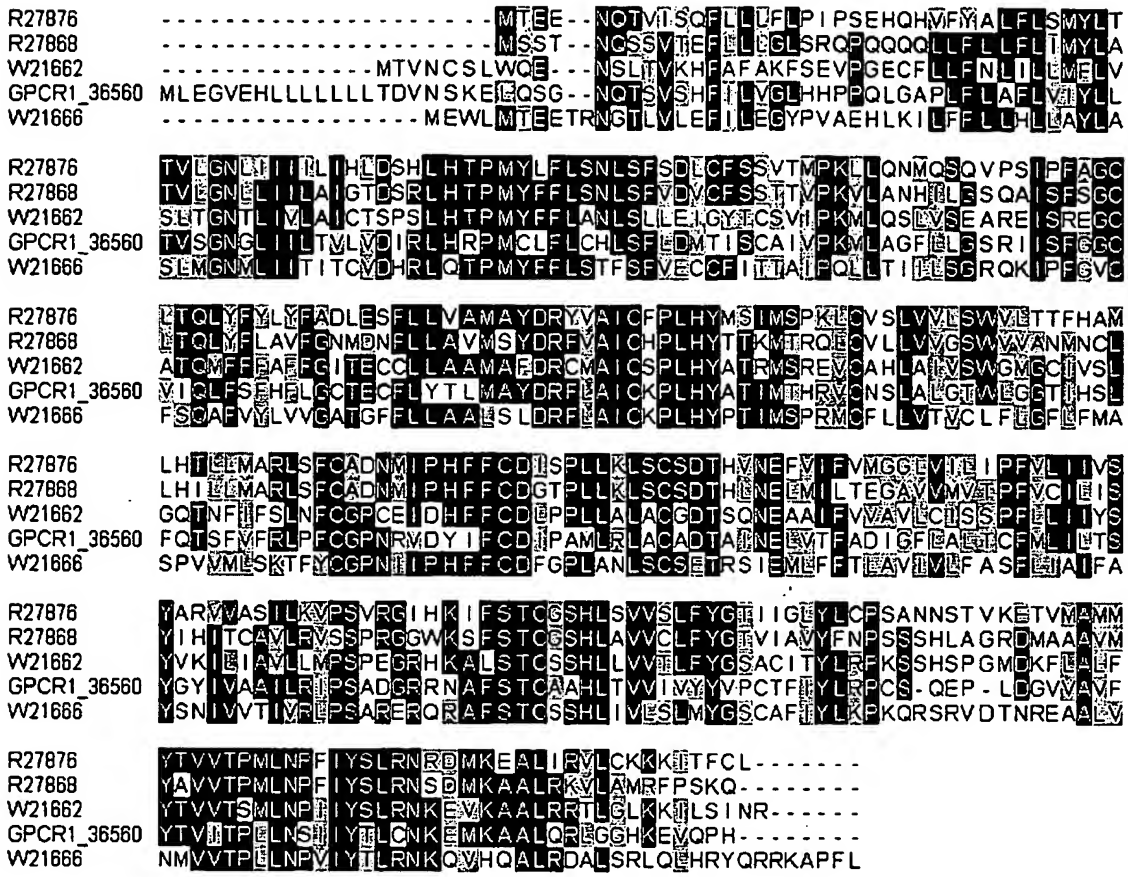
Identities = 185/297 (62%), Positives = 225/297 (75%), Frame = -3

```
Query:   980 TSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTVLVDIRLH-RPMCLFLCH 804
             T V+ F+L+GL HPP L   LFL FL+IY+LT  GN LI+LTV  D +LH RPM + L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   803 LSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMAYDRFLA 624
             LSFLDM +S  IVP+++  F   ++ I+FGGCV QL+ FHFLG T+CFLYTLMAYDR+LA
Sbjct:    71 LSFLDMWLSSVIVPRIILNFTPANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYLA 130

Query:   623 ICKPLHYATIMTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCDIPAML 444
             IC+PL Y  +M  ++C  L  G W+ G+IH   Q +  FRLP+CGP  VDY FCDIPA+L
Sbjct:   131 ICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAVL 190

Query:   443 RLACADTAINELVTFADIGFLALTCFMLILTSYGYIVAAILRIPSADGRRNAFSTCAAHL 264
             RLACADTAINELVTF DIG +A +CF+LIL SY  IV AIL+I +ADGRR AFSTC +HL
```

```
Sbjct:   191 RLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSHL 250

Query:   263 TVVIVYYVPCTFIYLRPCSQEPLDGVVAVFYTVITPLLNSIIYTLCNKEMKAALQRL 93 (SEQ
ID NO:69)
             TVV VYYVPC FIYLR  S+   DG VAVFYTV+TPLLN +IYTL N+E+ +AL+RL
Sbjct:   251 TVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRL 307 SEQ
ID NO:70
```

A multiple sequence alignment is given in Table 38, with the protein of the invention being shown on line 4, in a ClustalW analysis comparing the NOV5 protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV5 protein is a novel member of the OR family. In the following table, GPCR6_36560 is the full-length SEQ ID NO:10. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 38



Cellular localization analysis using Psort of the NOV5 protein of invention indicated that it might be targeted to the plasma membrane (Table 39).

TABLE 39

plasma membrane --- Certainty=0.6400(Affirmative) < succ>

Golgi body --- Certainty=0.4600(Affirmative) < succ>

endoplasmic reticulum (membrane) --- Certainty=0.3700(Affirmative) < succ>

endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>

SignalP analysis of the NOV5protein of invention indicated that it has got secretory signal peptide (Table 40). The first 70 amino acids of SEQ ID NO:10 (GPCR1_36560 (332 aa)) were used for signal peptide prediction.

TABLE 40

```
< Is the sequence a signal peptide?

# Measure  Position  Value  Cutoff  Conclusion

max. C    20        0.829  0.37    YES

max. Y    20        0.705  0.34    YES

max. S    57        0.968  0.88    YES

mean S    1-19      0.743  0.48    YES

# Most likely cleavage site between pos. 19 and 20: VNS-KE
```

The results of a search of the sequence databases using BLASTX is shown in Table 41. These results illustrate a very small P value when comparing the homology of NOV5 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV5 is a novel protein member of the GPCR family.

TABLE 41

```
Query= SEQ ID NO:9 (1120 letters)
 Translating both strands of query sequence in all 6 reading frames
Database: /opt/database/licensed/blast/geneseq_aa
354,275 sequences; 52,135,959 total letters.
Smallest
Sum
Reading High Probability
Sequences producing High-scoring Segment Pairs:      Frame Score P(N)     N
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ... +3  942 6.4e-94  1
patp:Y90874 Human G protein-coupled receptor GTAR14-5 ... +3  942 6.4e-94  1
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ... +3  922 8.4e-92  1
patp:Y90873 Human G protein-coupled receptor GTAR14-3 ... +3  922 8.4e-92  1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ... +3  772 6.6e-76  1
patp:Y90872 Human G protein-coupled receptor GTAR14-1 ... +3  772 6.6e-76  1

The highest extent of similarity in the above sequences is 60%.
```

NOV6

The novel nucleic acid of 1049 nucleotides (designated CuraGen Acc. No. AC019108_G) encoding a novel GPCR-like protein is shown in Table 42. An open reading frame was identified beginning with an atg initiation codon at nucleotides 28-30 and ending with a tag codon at nucleotides 961-963. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 42, and the start and stop codons are in bold letters. The encoded protein having 311 amino acid residues is presented using the one-letter code in Table 43.

TABLE 42

AGGGAGAGAGACCAAGGGTGAGAAGAAATGTCCAACGCCAGCCTACTGACAGCGTTCATCCTCATGGGCCTTCCCCAT
GCCCCAGCGCTGGACGCCCCCCTCTTTGGAGTCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACCTCCTCATC
CTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCACCATGTACTACTTCCTCACCAACCTGTCGTTCATTGACATG
TGGTTCTCCACTGTCACGGTGCCCAAATTGCTGATGACTTTGGTGTTCCCAAGTGGCAGGGCTATCTCCTTCCACAGC
TGCATGGCTCAGCTCTATTTCTTTCACTTCCTAGGGGCACCGACGTGTTTCCTCTACAGGGTCATGTCCTGTGATCGC
TACCTGGCCATCAGTTACCCGCTCAGGTACACCAGCATGATGACTGGGCGCTCGTGTACTCTTCTGGCCACCAGCACT
TGGCTCAGTGGCTCTCTGCACTCTGCTGTCCAGGCCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACTGGATC
CAGCACTATTTGTGTGATGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAGACTGTCATTTTT
GTGACTGTTGGAATAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGTGTCCATCGTCTGTTCCATCCTG
CGGATCCGCACCTCAGAGGGGAAGCACAGAGCCTTTCAGACCTGTGCCTCCCACTGTATCGTGGTCCTTTGCTTCTTT
GGCCCTGGTCTTTTCATTTACCTGAGGCCAGGCTCCAGGAAAGCTGTGGATGGAGTTGTGGCCGTTTTCTACACTGTG
CTGACGCCCCTTCTCAACCCTGTTGTGTACACCCTGAGGAACAAGGAGGTGAAGAAAGCTCTGTTGAAGCTGAAAGAC
AAAGTAGCACATTCTCAGAGCAAATAGACACTAGGGAAGATTACATATCTTAGCTCTTGTGAATAGTGCTGTGAAAAA
CATACAGGGGCAGGTATCTTTTGGA SEQ ID NO:11

TABLE 43

MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYFLTNLSFIDMWFS
TVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGAPTCFLYRVMSCDRYLAISYPLRYTSMMTGRSCTLLA
TSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPILKLACADTSAIETVIFVTVGIVASGCFVLIVLSYV
SIVCSILRIRTSEGKHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRN
KEVKKALLKLKDKVAHSQSX SEQ ID NO:12

In a search of sequence databases, it was found, for example, that the nucleic acid sequence (SEQ ID NO:11, QUERY) has 595 of 897 bases (66%) identical to a Rat species GPCR mRNA (GENBANK-ID: U5094, Subject9) (Table 44). The full amino acid sequence of the protein of the invention (SEQ ID NO:12, Query) was found to have 164 of 298 amino acid residues (55%) identical to, and 222 of 298 residues (74%) positive with, the 318 amino acid residue protein from Rat species (ptnr:SPTREMBL-ACC:Q62944, Subject) (Table 45).

TABLE 44

gb:GENBANK-ID:RNU50949|acc:U50949 Rattus norvegicus taste bud receptor protein TB 641 (TB 641) gene, complete cds - Rattus norvegicus, 1278 bp. Length = 1278

Minus Strand HSPs:

Score = 1499 (224.9 bits), Expect = 9.2e-62, P = 9.2e-62

Identities = 595/897 (66%), Positives = 595/897 (66%), Strand = Minus / Plus

```
Query:   993 TGACAGCGTTCATCCTCATGGGCCTTCCCCATGCCCCAGCGCTG-GACGCCCCCCTCTTT 935
             TGACAG  TTC T CTC TGGGC T  C CAT CCCCA   CT  GA  C   CCTCTT
Sbjct:   173 TGACAGATTTCCTTCTCCTGGGCTTGGCTCATCCCCCAAATCTAAGAA-CGTTCCTCTTC 231

Query:   934 GGAGTCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACCTCCTCATCCTGCTGGTG 875
                GTCTTCCT  T  TTTAC T CT AC   G TGGGGAACCT CTCATCCTGCT
Sbjct:   232 CTGGTCTTCCTCCTCATTTACATCCTGACACAGTTGGGGAACCTGCTCATCCTGCTCACA 291

Query:   874 ATCAGGGTGGATTCTCACCTCCA--CAC-CACCATGTACTACT-TCCTCACCAACCTGTC 819
              T  GGG  GA  C  A CT CA  C C C CCATGTAC A T T CT   C   CT TC
Sbjct:   292 GTGTGGGCTGACCCCAAGCTGCATGCCCGCCCCATGTAC-ATTCTGCTGGGCGTGCTCTC 350

Query:   818 GTTCATTGACATGTGGTTCTCCACTGTCACGGTGCC-CAAATTGCTGATGACTTTGGTGT 760
              TTC T GACATGTGG TCTCC C GTCA  GT CC C AATT  T    ACTT   T
Sbjct:   351 CTTCCTGGACATGTGGCTCTCCTCAGTCATTGTCCCTCGAATTATTTTAAACTTCACTCC 410

Query:   759 TCCCAAGTGGCAGGGCTATCTCCTTCCACAGCTGCATGGCTCAGCTCTATTTCTTTCACT 700
             T CCAA    CA GGCTATC C TT     GCTG  T GCTCA CTCTATTT TT CACT
Sbjct:   411 TGCCAA----CAAGGCTATCGCATTTGGTGGCTGTGTAGCTCAACTCTATTTTTTCCACT 466

Query:   699 TCCTAGGG-GCACCGACGTGTTTCCTCTACAGGGTCATGTCCTGTGATCGCTACCTGGCC 641
             TCCT GG  GCACC A GTG TTCCTCTA A   T ATG CCT TGA  G TACCTGGC
Sbjct:   467 TCCTGGGCAGCACCCA-GTGCTTCCTCTATACCTTGATGGCCTATGACAGGTACCTGGCA 525

Query:   640 ATCAGTTACCCGCTCAGGTACACCA-GCATGATGACTGGGCGCTCGTGTACTCTTCTGGC 582
             AT  GT A CC CT  G TAC C   GC T ATGA TGGG   T  TG AC  T CTGG
Sbjct:   526 ATATGTCAGCCTCTTCGCTACCCTGTGC-TCATGAATGGGAAGTTATGCACAATCCTGGT 584

Query:   581 CACCAGCACTTGGCTCAGTGGCTCTCTGCACTCTG-CTGTCCAGGCCATATTGACTTTCC 523
               C  G  CTTGG T   TGGCTC  T CA T  G CT T CA GCCA   TGAC TTCC
Sbjct:   585 GGCTGGAGCTTGGGTGGCTGGCTCCATCCA-TGGGTCTATTCAAGCCACTCTGACCTTCC 643

Query:   522 ATTTGCCCTACTGTGGACCCAACTGGA-TCCAGCACTATTTGTGTGATGCACCGCCCATC 464
               TTGCCCTACTGTGG CC AA  G A T  A  ACT  TT TGTGA    CC  C  T
Sbjct:   644 GATTGCCCTACTGTGGGCCTAAG-GAAGTGGATTACTTCTTCTGTGACATTCCTGCAGTG 702

Query:   463 CTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAGACTG-TCATTTTTGTG-ACTGTTG 406
             CTGA ACTGGCCTGTGC GA AC  CA  CA  GA ACTG T A  TTTGTG AC  TTG
Sbjct:   703 CTGAGACTGGCCTGTGCTGATACAGCAATCAATGA-ACTGGTGACCTTTGTGGACA-TTG 760

Query:   405 GAATAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGTGTCCATCGTCTGTT 346
             G  TAGTGGC  C  G TGCTT  T CTGAT  TGCT TCCTA G    CAT GT   T
Sbjct:   761 GGGTAGTGGCTGCCAGTTGCTTCCTGCTGATTCTGCTCTCCTACGCCAACATAGTTCATG 820

Query:   345 CCATCCTGCGGATCCGCACCTCAGAGGGGAAGC-ACAGAGCCTT-TCAGACCTGTGCCTC 288
             CCATCCTG  GAT CGCAC  CAGA GG A G  AC G GCCTT TC  ACCTGTG CTC
Sbjct:   821 CCATCCTGAAGATACGCACTGCAGATGGCAGGAGAC-GTGCCTTCTCC-ACCTGTGGCTC 878
```

```
Query:   287 CCACTGTATC-GTGGTC-CTTTGCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGCCA 230
             CCA T T  C GTGGTC C  T CT CT TG CCC  GT TTTTCAT TACCT  GG CA
Sbjct:   879 CCA-TCTCACTGTGGTCACAGT-CTACTATGTCCCCTGTATTTTCATCTACCTTCGGGCA 936

Query:   229 GGCTCCAGGAAAGCTGTGGATGGAGTTGTGGCCGTTTTCTACACTGTGCTGACGCCCCTT 170
             GG TCCA GA   C  T GA GGAG  GT GC GT TT TACACTGT  T AC CC  T
Sbjct:   937 GGTTCCAAGAGTTCCTTTGACGGAGCAGTTGCTGTATTTTACACTGTTGTCACTCCATTA 996

Query:   169 CTCAACCCTGTTGTGTACACCCTGAGGAACAAGGAGGTGAAGAAAGCTCTGTTGAAGCTG 110
             CT AA CC  T  T TACAC CTGAGGAAC AGGA GTGAA    GC CTG  GA GCT
Sbjct:   997 CTGAATCCCCTCATCTACACTCTGAGGAACCAGGAAGTGAATTCTGCCCTGAAGAGGCTA 1056

Query:   109 AAAGACAAAGTAG 97  SEQ ID NO:75
             A AG CA  GTAG
Sbjct:  1057 AGAG-CAG-GTAG 1067   SEQ ID NO:76
```

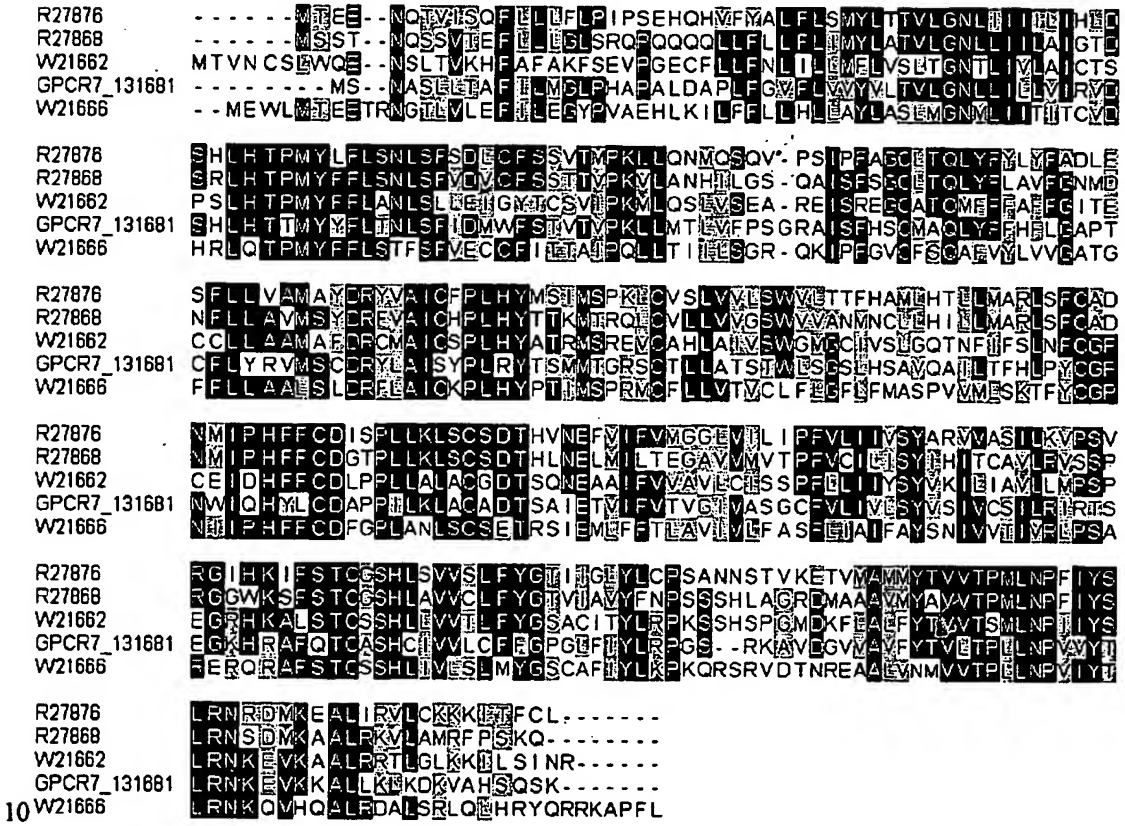
TABLE 45

```
ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - RATTUS NORVEGICUS
(RAT), 318 aa.  Length = 318
  Minus Strand HSPs:
 Score = 895 (315.1 bits), Expect = 6.7e-89, P = 6.7e-89
 Identities = 164/298 (55%), Positives = 222/298 (74%), Frame = -1

Query:  1000 SLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTT-MYYFLTN 824
              +++T F+L+GL H P L   LF VFL++Y+LT LGNLLILL +  D  LH   MY  L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   823 LSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGAPTCFLYRVMSCDRYL 644
             LSF+DMW S+V VP++++    P+ +AI+F  C+AQLYFFHFLG+  CFLY +M+ DRYL
Sbjct:    71 LSFLDMWLSSVIVPRIILNFT-PANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYL 129

Query:   643 AISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPI 464
             AI  PLRY  +M G+ CT+L    W++GS+H ++QA LTF LPYCGP  + ++ CD P +
Sbjct:   130 AICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAV 189

Query:   463 LKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTCASH 284
             L+LACADT+  E V FV +G+VA+ CF+LI+LSY +IV +IL+IRT++G+ RAF TC SH
Sbjct:   190 LRLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSH 249

Query:   283 CIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNKEVKKALLKLK 107
SEQ ID NO:77
                VV  ++ P +FIYLR GS+ + DG VAVFYTV+TPLLNP++YTLRN+EV  AL +L+
Sbjct:   250 LTVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRLR 308
SEQ ID NO:78
```

A multiple sequence alignment is given in Table 46, with the protein of the invention being shown on line 4, in a ClustalW analysis comparing the NOV6 protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric

and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV6 protein is a novel member of the OR family. In the following table, GPCR7_131681 is the full-length SEQ ID NO:12. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 46



Cellular localization analysis using Psort of the NOV6 protein of invention indicated that it might be targeted to the plasma membrane (Table 47).

TABLE 47

```
plasma membrane --- Certainty=0.6400(Affirmative) < succ>
Golgi body --- Certainty=0.4600(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3700(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>
```

SignalP analysis of the protein of invention indicated that it has got secretory signal peptide (Table 48). The first 70 amino acids of SEQ ID NO:12 (GPCR7_131681 (311 aa)) were used for signal peptide prediction.

These results illustrate a very small P value when comparing the homology of NOV6 to various members of the GPCR family. Such degree of homology, based on the small P values, is very unlikely to have occurred by chance alone. Accordingly, NOV6 is a novel protein member of the GPCR family.

TABLE 48

```
< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C    52        0.702  0.37    YES
  max. Y    52        0.617  0.34    YES
  max. S    36        0.980  0.88    YES
  mean S     1-51     0.825  0.48    YES
# Most likely cleavage site between pos. 51 and 52: VDS-HL
```

The results of a search for homology of SEQ ID NO:11 in the sequence databases using BLASTX is shown in Table 49

TABLE 49

QUERY= AC019108_G CURA_150 GPCR
(1039 LETTERS)

TRANSLATING BOTH STRANDS OF QUERY SEQUENCE IN ALL 6 READING FRAMES

DATABASE: /OPT/DATABASE/LICENSED/BLAST/GENESEQ_AA
354,275 SEQUENCES; 52,135,959 TOTAL LETTERS.

SEQUENCES PRODUCING HIGH-SCORING SEGMENT PAIRS:	READING FRAME	SMALLEST SUM HIGH SCORE	PROBABILITY P(N)	N
PATP:Y90874 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-5 ...	+1	880	2.4E-87	1
PATP:Y90874 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-5 ...	+1	880	2.4E-87	1
PATP:Y90873 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-3 ...	+1	815	1.8E-80	1
PATP:Y90873 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-3 ...	+1	815	1.8E-80	1
PATP:Y90872 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-1 ...	+1	705	8.3E-69	1
PATP:Y90872 HUMAN G PROTEIN-COUPLED RECEPTOR GTAR14-1 ...	+1	705	8.3E-69	1

THE HIGHEST EXTENT OF SIMILARITY IN THE ABOVE SEQUENCES IS 55%.

NOV7

In the present invention, the target sequence identified previously, NOV6 (Accession Number AC019108_G), was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a library containing a wide range of cDNA species. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated NOV7 (Accession Number AC019108D). NOV7 is 1039 nucleotides encoding a novel GPCR-like protein is shown in Table 42. An open reading frame was identified beginning with an atg initiation codon at nucleotides 28-30 and ending with a tag codon at nucleotides 960-962. The encoded protein having 311 amino acid residues is presented using the one-letter code in Table 51.

TABLE 50

AGGGAGAGAGACCAAGGGTGAGAAGAAATGTCCAACGCCAGCCTACTGACAGCGTTCATCCTCACGGGCCTTCCCCATGC
CCCAGCGCTGGACGCCCCCCTCTTTGGAGTCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACCTCCTCATCCTGC
TGGTGATCAGGGTGGATTCTCACCTCCACACCACCATGTACTACTTCCTCACCAACCTGTCGTTCATTGACATGTGGTTC
TCCACTGTCACGGTGCCCAAATTGCTGATGACTTTGGTGTTCCCAAGTGGCAGGACTATCTCCTTCCACAGCTGCATGGC
TCAGCTCTATTTCTTTCACTTCCTAGGGGGCACCGAGTGTTTCCTCTACACAGTCATGTCCTGTGATCGCTACCTGGCCA

TCAGTTACCCGCTCAGGTACACCAGCATGATGACTGGGCGCTCGTGTACTCTTCTGGCCACCAGCACTTGGCTCAGTGGC
TCTCTGCACTCTGCTGTCCAGGCCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACTGGATCCAGCACTATTTGTG
TGATGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAGACTGTCATTTTTGTGACTGTTGGAATAG
TGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGTGTCCATCGTCTGTTCCATCCTGCGGATCCGCACCTCAGAG
GGGAAGCACAGAGCCTTTCAGACCTGTGCCTCCCACTGTATCGTGGTCCTTTGCTTCTTTGGCCCTGGTCTTTTCATTTA
CCTGAGGCCAGGCTCCAGGAAAGCTGTGGATGGAGTTGTGGCCGTTTTCTACACTGTGCTGACGCCCCTTCTAAACCCTG
TTGTGTACACCCTGAGGAACAAGGAGGTGAAGAAAGCTCTGTTGAAGCTGAAAGACAAAGTAGCACATTCTCAGAGCAAA
TAGACACTAGGGAAGATTACATATCTTAGCTCTTGTGAATAGTGCTGTGAAA SEQ ID NO:13

TABLE 51

MSNASLLTAFILTGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYFLTNLSFIDMWFSTV
TVPKLLMTLVFPSGRTISFHSCMAQLYFFHFLGGTECFLYTVMSCDRYLAISYPLRYTSMMTGRSCTLLAT
STWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPILKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVS
IVCSILRIRTSEGKHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNK
EVKKALLKLKDKVAHSQSK SEQ ID NO:14

Cellular localization analysis using Psort of the protein of invention indicated that NOV7t might be targeted to the plasma membrane (certainity =0.6400). The NOV7 polypeptide seems to have a cleavable N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 51 and 52.

The BLASTP hits are given in Table 52 (general) and Table 53 (human). These results illustrate a very small P value when comparing the homology of NOV7 to various members of the GPCR family. Such degree of homology based on the P values indicates that NOV7 is a novel member of the GPCR family.

TABLE 52

Score = 902 (317.5 bits), Expect = 2.6e-90, P = 2.6e-90 Identities = 165/298 (55%), Positives = 222/298 (74%) with ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus norvegicus (Rat), 318 aa.

Score = 743 (261.5 bits), Expect = 1.8e-73, P = 1.8e-73 Identities = 145/302 (48%), Positives = 201/302 (66%) with ACC:Q9R0K3 ODORANT RECEPTOR MOR83 - Mus musculus (Mouse), 308 aa.

Score = 739 (260.1 bits), Expect = 4.8e-73, P = 4.8e-73 Identities = 143/299 (47%), Positives = 201/299 (67%) with ACC:Q9R0K4 ODORANT RECEPTOR MOR10 - Mus musculus (Mouse), 310 aa.

TABLE 53

Score = 692 (243.6 bits), Expect = 4.6e-68, P = 4.6e-68 Identities = 133/303 (43%), Positives = 200/303 (66%) with ACC:Q9UGF6 BA150A6.2 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN (HS6M1-21)) - Homo sapiens (Human), 321 aa.

NOV8

The novel nucleic acid of 985 nucleotides (SEQ ID NO:15) (designated CuraGen Acc. No. CG50373-01) encoding a novel Olfactory Receptor-like protein is shown in Table 54. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 9-11 and ending with a TGA codon at nucleotides 978-980. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 323 amino acid residues is presented using the one-letter code in Table 55.

TABLE 54

```
TACTGAAGATGAACCCTGAAAACTGGACTCAGGTAACAAGCTTTGTCCTTCTGGGTTTCC    60
CCAGTAGCCACCTCATACAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAA   120
CAGCCACAGGCAAGCTGCTAATTATTGTGCTCAGCTGGATAGACCAACGCCTGCACATAC   180
AGATGTACTTCTTCCTGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGG   240
TTCCCAAGATGCTTGTCGTCATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCA   300
TCATCCAGTCCTACCTCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCTTGGCCGTCA   360
TGTCTCTGGATCGTTACCCGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATG   420
GCCATGTCTCTTCCCAACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCTCTGGGTCCTTT   480
GCCCCACTGTCCTCATGGCCAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCT   540
TTCGTGACAGTTGGCCCTTGCTCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGG   600
CGGCTTTCATGCTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACCTCAGTTTCCT   660
ATGCCTGCATTCTTGCCACTGTTCTCAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGT   720
TTTCCACTTGTGCCTCGCATCTTACAGTGGTGGTCATCATCTATGGCAGTTCCATCTTTC   780
TCTACATTCGTATGTCAGAGGCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTCCTGA   840
GCTGCATCATCACACCCTTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGC   900
AGCAAGCACTGAGAGAAGCCTTGGGGTGGCCCAGGCTCACTGCTGTGATGAAACAGAGGG   960
TCACAAGTCAAAGGAAATGATCTTA SEQ ID NO:15                          985
```

TABLE 55

```
MNPENWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMY   60
FFLRNFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSL   120
DRYPAICRPLRYETLMNGHVSSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRD   180
SWPLLRLSCGDTHLLKLAAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFST   240
CASHLTVVVIIYGSSIFLYIRMSEAQSKLLNKGASVLSCIITPFLNPFIFTLRNDKVQQA   300
LREALGWPRLTAVMKQRVTSQRK SEQ ID NO:16                            323
```

In a search of sequence databases, it was found, for example, that the nucleic acid sequence of this invention (SEQ ID NO:15, Query) has 552 of 900 bases (61%) identical to a gb:GENBANK-ID:AF102523|acc:AF102523.1 mRNA from Mus musculus (Mus musculus olfactory receptor C6 gene, complete cds, Subject) (Table 56 and 58). The full amino acid sequence of the protein of the invention (SEQ ID NO:16, Query) was found to have 151 of 300 amino acid residues (50%) identical to, and 213 of 300 amino acid residues (71%) similar to, the

313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus musculus (Subject) (Mouse) (OLFACTORY RECEPTOR C6)(Table 57 and 58).

TABLE 56

```
>gb:GENBANK-ID:AF102523|acc:AF102523.1 Mus musculus olfactory receptor C6
gene, complete cds - Mus musculus, 942 bp.
            Length = 942

  Plus Strand HSPs:

 Score = 1142 (171.3 bits), Expect = 1.3e-45, P = 1.3e-45
 Identities = 552/900 (61%), Positives = 552/900 (61%), Strand = Plus / Plus

Query:    21 AACTGGACTCAGGTAACA-AGCTTTGTCCTTCTGGGTTTCCCCAGTAGCCACCTCA--TA 77
             AAC G ACT   GT AC  AG TTT T  T CTGGG  T  C AG  GCC   T A  T
Sbjct:     7 AACAGCACTACTGTTACTGAG-TTTATTTTGCTGGGGCTGTC-AGATGCCTG-TGAGCTG 63

Query:    78 CAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAACAGCCACAGGCAAGCTG 137
             CAG T CT  T TTCCTGGG TT  T  TGACCTAC T  T A   C  C GG AA CT
Sbjct:    64 CAGGTGCTCATATTCCTGGGCTTTCTCCTGACCTACTTCCTCATT-CTGCTGGGAAACTT 122

Query:   138 C-TAATTATTGTGCTCAGC-TGGATAGACCAACGCCTGCACATACAGATGTACTTCTTCC 195
             C T AT AT  T  TCA C T G T GAC   CGCCT  ACA  C  ATGTA T CTTCC
Sbjct:   123 CCTCATCATCTTCATCACCCTTG-TGGACAGGCGCCTTTACACCCCCATGTATTACTTCC 181

Query:   196 TGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGGTTCCCAAGATGCTTG 255
             T CG AA TT  CC T CTGGAG T T G T     CTGT  T  T CCCAAGATGCT
Sbjct:   182 TCCGCAACTTTGCCATGCTGGAGATCTGGTTCACCTCTGTCATCTTCCCCAAGATGCTAA 241

Query:   256 TCGTCATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCATCATCCAGTCCTACC 315
               C  CATC TCAC GG  AT A ACCATCTC  T  T  G TG  TC TCCA  C T CC
Sbjct:   242 CCAACATCATCACAGGACATAAGACCATCTCCCTACTAGGTTGTTTCCTCCAAGCATTCC 301

Query:   316 TCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCTTGGCCGTCATGTCTCTGGATCGTT 375
             TCTA TTCTT CT GGCACCACTGA TTCTT CT  TGGC GT ATGTC  T GA  G T
Sbjct:   302 TCTATTTCTTCCTTGGCACCACTGAGTTCTTTCTACTGGCAGTGATGTCCTTTGACAGGT 361

Query:   376 ACCCGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATGGCCATGTCTCTTCCC 435
             A   GGC AT TG    CC  T CG TATG  ACC T ATGA       GTCT T  CC
Sbjct:   362 ATGTGGCCATTTGTAACCCTTTGCGTTATGCCACCATTATGAGCAAAAGAGTCTGTGTCC 421

Query:   436 AACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCT-CTGGGTCCTTTGCCCCACTGTCC-T 493
             A CT GTG T   CTC TGG T  CTGGATT CT CT   TC T    CC A T TC  T
Sbjct:   422 AGCTTGTGTTTTGCTCATGGATGTCTGGATTGCTTCTCA-TCATAGTTCCTAGT-TCAAT 479

Query:   494 CATGGC-CAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCTTTCGTGACAGTT 552
               T    CAGC  GCC TTCTGTGGCCC AA    ATT A CA TTCTT  GTGACA  T
Sbjct:   480 TGTATTTCAGCA-GCCATTCTGTGGCCCAAACATCATTAATCATTTCTTCTGTGACAACT 538

Query:   553 GGCC-CTTGCTCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGGC-GGCTTTCAT 610
               CC CTT    A  CT    TGTG  GA AC   CCTG T A A T  C GG TTT  T
Sbjct:   539 TTCCACTTATGGAA-CTCATATGTGCAGATACTAGCCTGGT-AGAGTTCCTGGGTTTTGT 596

Query:   611 GCTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACCTCAGTTTCCTATGCCTGCAT 670
               T  C A  TT     T CTGGGC C CTGGCT TGAC  C    T CTATG C  CAT
```

```
Sbjct:   597 TATTGCCAATTTCAGCCTCCTGGGCACTCTGGCTGTGACTGCCACCTGCTATGGCCACAT 656

Query:   671 TCTTGCCACTGTTCT-CAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGTTTTCCACTT 729
             TCT    AC  TTCT CA    CC T CAGC    GAG G A GAAAGC TT TC ACTT
Sbjct:   657 TCTCTATACCATTCTACACATTCCTT-CAGCCAAGGAGAGGAAGAAAGCCTTCTCAACTT 715

Query:   730 GTGCCTCGCATCTTACAGTGGTGG-TCATCATCTATGGCAGTTCCATCTTTCTCTACATT 788
             G  CCTC CAT TTA  GTGGTG  TC TC TCTA GGCAG T  ATCTT  T TA  T
Sbjct:   716 GCTCCTCTCATATTATTGTGGTGTCTC-TCTTCTACGGCAGCTGTATCTTCATGTATGTC 774

Query:   789 CG-TATGTCA-GAG-GCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTC-CTGAGCTG 844
             CG T TG CA GA  G  CAG    A    C  AACAA GGTG    C T  CT A C
Sbjct:   775 CGGTCTGGCAAGAATGGACAGGGGGAGGATCATAACAA-GGTGGTGGCATTGCTCAACAC 833

Query:   845 CATCATCACACCCTTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGCAGCA 904
               T  T ACACCC    T AACCC TTCATCT CACTCT  G AA  A  AGGTG AGCA
Sbjct:   834 TGTAGTGACACCCACACTCAACCCCTTCATCTACACTCTGAGGAACAAGCAGGTGAAGCA 893

Query:   905 AGCACTGAGAGAA 917  SEQ ID NO:83
             G A T AG GAA
Sbjct:   894 GGTATTTAGGGAA 906  SEQ ID NO:84
```

TABLE 57

```
>ptnr:SPTREMBL-ACC:Q9Z1V0 OLFACTORY RECEPTOR C6 - Mus musculus (Mouse), 313
aa.
            Length = 313

 Score = 810 (285.1 bits), Expect = 1.7e-80, P = 1.7e-80
 Identities = 151/300 (50%), Positives = 213/300 (71%)

Query:     5 NWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFLR 64
             N T VT F+LLG   +  +Q L+FLG ++TY +   G  LII ++ +D+RL+  MY+FLR
Sbjct:     3 NSTTVTEFILLGLSDACELQVLIFLGFLLTYFLILLGNFLIIFITLVDRRLYTPMYYFLR 62

Query:    65 NFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRYP 124
             NF+ LE+   +V+ PKML  I+TG  TIS + C +Q++LYFFLGTT+FFLLAVMS DRY
Sbjct:    63 NFAMLEIWFTSVIFPKMLTNIITGHKTISLLGCFLQAFLYFFLGTTEFFLLAVMSFDRYV 122

Query:   125 AICRPLRYETLMNGHVSSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRDSWPL 184
             AIC PLRY T+M+  V  QLV  SW++G L ++ P+ ++    PFCGPN I+HFF D++PL
Sbjct:   123 AICNPLRYATIMSKRVCVQLVFCSWMSGLLLIIVPSSIVFQQPFCGPNIINHFFCDNFPL 182

Query:   185 LRLSCGDTHLLKLAAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFSTCASH 244
             + L C DT L++   F+++   LLG+LA+T+  Y  IL T+L  P+A ER+KAFSTC+SH
Sbjct:   183 MELICADTSLVEFLGFVIANFSLLGTLAVTATCYGHILYTILHIPSAKERKKAFSTCSSH 242

Query:   245 LTVVVIIYGSSIFLYIRMSE-AQSKLLNKGASVLSCIITPFLNPFIFTLRNDKVQQALRE 303
SEQ ID NO:85
             + VV + YGS IF+Y+R  +  Q +  NK  ++L+ ++TP LNPFI+TLRN +V+Q  RE
Sbjct:   243 IIVVSLFYGSCIFMYVRSGKNGQGEDHNKVVALLNTVVTPTLNPFIYTLRNKQVKQVFRE 302
SEQ ID NO:86
```

TABLE 58

```
>s3aq:137032954 Category C: 40 frag (40 non-5'sig-CG), 985 bp.
          Length = 985

  Plus Strand HSPs:

 Score = 4916 (737.6 bits), Expect = 7.6e-217, P = 7.6e-217
 Identities = 984/985 (99%), Positives = 984/985 (99%), Strand = Plus / Plus

Query:     1 TACTGAAGATGAACCCTGAAAACTGGACTCAGGTAACAAGCTTTGTCCTTCTGGGTTTCC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 TACTGAAGATGAACCCTGAAAACTGGACTCAGGTAACAAGCTTTGTCCTTCTGGGTTTCC 60

Query:    61 CCAGTAGCCACCTCATACAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 CCAGTAGCCACCTCATACAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAA 120

Query:   121 CAGCCACAGGCAAGCTGCTAATTATTGTGCTCAGCTGGATAGACCAACGCCTGCACATAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 CAGCCACAGGCAAGCTGCTAATTATTGTGCTCAGCTGGATAGACCAACGCCTGCACATAC 180

Query:   181 AGATGTACTTCTTCCTGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   181 AGATGTACTTCTTCCTGCGGAATTTCTCCTTCCTGGAGCTGTTGCTGGTAACTGTTGTGG 240

Query:   241 TTCCCAAGATGCTTGTCGTCATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCA 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   241 TTCCCAAGATGCTTGTCGTCATCCTCACGGGGGATCACACCATCTCATTTGTCAGCTGCA 300

Query:   301 TCATCCAGTCCTACCTCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCTTGGCCGTCA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   301 TCATCCAGTCCTACCTCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCTTGGCCGTCA 360

Query:   361 TGTCTCTGGATCGTTACCCGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   361 TGTCTCTGGATCGTTACCCGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATG 420

Query:   421 GCCATGTCTCTTCCCAACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCTCTGGGTCCTTT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   421 GCCATGTCTCTTCCCAACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCTCTGGGTCCTTT 480

Query:   481 GCCCCACTGTCCTCATGGCCAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCT 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   481 GCCCCACTGTCCTCATGGCCAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCT 540

Query:   541 TTCGTGACAGTTGGCCCTTGCTCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   541 TTCGTGACAGTTGGCCCTTGCTCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGG 600

Query:   601 CGGCTTTCATGCTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACCTCAGTTTCCT 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   601 CGGCTTTCATGCTCTCTACGTTGGTGTTACTGGGCTCACTGGCTCTGACCTCAGTTTCCT 660

Query:   661 ATGCCTGCATTCTTGCCACTGTTCTCAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   661 ATGCCTGCATTCTTGCCACTGTTCTCAGGGCCCCTACAGCTGCTGAGCGAAGGAAAGCGT 720
```

```
Query:   721 TTTCCACTTGTGCCTCGCATCTTACAGTGGTGGTCATCATCTATGGCAGTTCCATCTTTC 780
             ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   721 TTTCCACTTGCGCCTCGCATCTTACAGTGGTGGTCATCATCTATGGCAGTTCCATCTTTC 780

Query:   781 TCTACATTCGTATGTCAGAGGCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTCCTGA 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   781 TCTACATTCGTATGTCAGAGGCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTCCTGA 840

Query:   841 GCTGCATCATCACACCCTTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   841 GCTGCATCATCACACCCTTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGC 900

Query:   901 AGCAAGCACTGAGAGAAGCCTTGGGGTGGCCCAGGCTCACTGCTGTGATGAAACAGAGGG 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   901 AGCAAGCACTGAGAGAAGCCTTGGGGTGGCCCAGGCTCACTGCTGTGATGAAACAGAGGG 960

Query:   961 TCACAAGTCAAAGGAAATGATCTTA 985  SEQ ID NO:87
             |||||||||||||||||||||||||
Sbjct:   961 TCACAAGTCAAAGGAAATGATCTTA 985  SEQ ID NO:88
```

A multiple sequence alignment is given in Table 59A and B, with the NOV8 protein of the invention (SEQ ID NO:16) being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences. In the alignment shown below, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV8 protein is a novel member of the OR family. In the following table, CG50373-01 is the full-length SEQ ID NO:16. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 59A

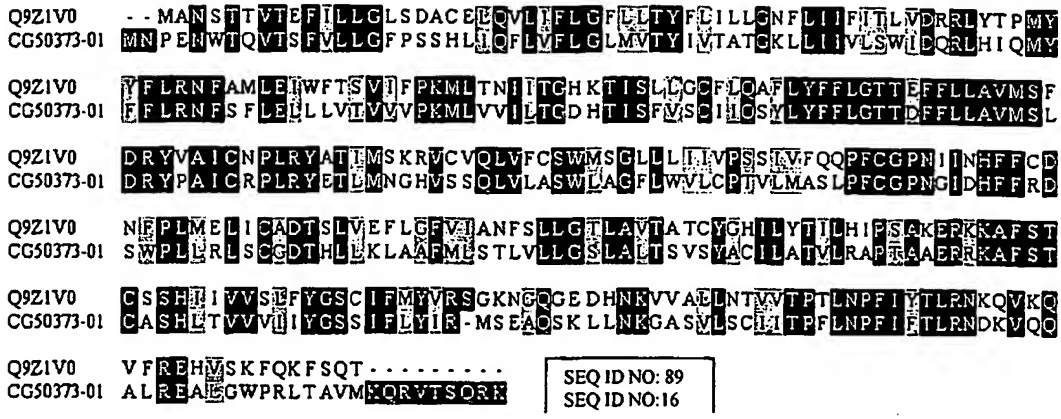


TABLE 59B

Accno	Common Name	Length
CG50373-01	novel Olfactory Receptor-like protein	
Q9Z1V0	OLFACTORY RECEPTOR C6	313

The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name 7tm_1 (7 transmembrane receptor (rhodopsin family)) at amino acid positions 41 to 290. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Tissue expression

NOV8 is expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue. Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen,

stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs. "Depth" rerepresents the number of clones covering the region of the SNP. The Putative Allele Frequency (Putative Allele Freq.) is the fraction of all the clones containing the SNP. The sign ">" means "is changed to".

Cons.Pos.: 65 Depth: 26 Change: T > C
Putative Allele Freq.: 0.192
Cons.Pos.: 76 Depth: 26 Change: T > C
Putative Allele Freq.: 0.077
Cons.Pos.: 106 Depth: 26 Change: T > G
Putative Allele Freq.: 0.115
Cons.Pos.: 199 Depth: 26 Change: C > T
Putative Allele Freq.: 0.077
Cons.Pos.: 294 Depth: 25 Change: C > T
Putative Allele Freq.: 0.080
Cons.Pos.: 431 Depth: 22 Change: C > G
Putative Allele Freq.: 0.364
Cons.Pos.: 615 Depth: 11 Change: C > G
Putative Allele Freq.: 0.273
Cons.Pos.: 732 Depth: 17 Change: C > T
Putative Allele Freq.: 0.353
Cons.Pos.: 760 Depth: 18 Change: A > G
Putative Allele Freq.: 0.111
Cons.Pos.: 859 Depth: 18 Change: C > T
Putative Allele Freq.: 0.222
Cons.Pos.: 956 Depth: 18 Change: T > A
Putative Allele Freq.: 0.222

Cellular Localization and Sorting

The Psort and Hydropathy profile for NOV8 is shown in Table 60. The results predict that NOV8 has a signal peptide and is likely to be localized on the plasma membrane with a certainty of 0.6400. The first 41 amino acids are more likely to be cleaved as a signal peptide based on the SignalP result (Table 60).

TABLE 60

plasma membrane --- Certainty=0.6400(Affirmative) < succ>

Golgi body --- Certainty=0.4600(Affirmative) < succ>

endoplasmic reticulum (membrane) --- Certainty=0.3700(Affirmative) < succ>

endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>

INTEGRAL Likelihood = -8.92 Transmembrane 71 - 87 (66 - 89)

INTEGRAL Likelihood = -6.00 Transmembrane 197 - 213 (192 - 217)

INTEGRAL Likelihood = -3.66 Transmembrane 150 - 166 (142 - 170)

INTEGRAL Likelihood = -2.97 Transmembrane 244 - 260 (241 - 263)

INTEGRAL Likelihood = -2.23 Transmembrane 276 - 292 (276 - 292)

INTEGRAL Likelihood = -0.22 Transmembrane 92 - 108 (91 - 108)

Likely a Type IIIa membrane protein (clv)

Is the sequence a signal peptide?

# Measure	Position	Value	Cutoff	Conclusion
max. C	39	0.534	0.37	YES
max. Y	42	0.455	0.34	YES
max. S	33	0.965	0.88	YES
mean S	1-41	0.756	0.48	YES

Most likely cleavage site between pos. 41 and 42: ATG-KL

NOV9

In the present invention, the target sequence identified previously, NOV4, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of

human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, NOV9, which is designated Accession Number AC019108E_da1. This differs from the previously identified sequence (Accession Number AC019108_E) at bp: 378, 429, 600, 857, 954.

NOV9 has 983 nucleotides and encodes a novel Olfactory Receptor-like protein as is shown in Table 61. An open reading frame was identified beginning with an ATG initiation codon at nucleotide 8 and ending with a TGA codon at nucleotides 977. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein is presented using the one-letter code in Table 62.

TABLE 61

ACTGAAGATGAACCCTGAAAACTGGACTCAGGTAACAAGCTTTGTCCTTCTGGGTTTCCCCAGTAGCCAC
CTCATACAGTTCCTGGTGTTCCTGGGGTTAATGGTGACCTACATTGTAACAGCCACAGGCAAGCTGCTAA
TTATTGTGCTCAGCTGGATAGACCAACGCCTGCACATACAGATGTACTTCTTCCTGCGGAATTTCTCCTT
CCTGGAGCTGTTGCTGGTAACTGTTGTGGTTCCCAAGATGCTTGTCGTCATCCTCACGGGGGATCACACC
ATCTCATTTGTCAGCTGCATCATCCAGTCCTACCTCTACTTCTTTCTAGGCACCACTGACTTCTTCCTCT
TGGCCGTCATGTCTCTGGATCGTTACCCGGCAATCTGCCGACCACTCCGCTATGAGACCCTGATGAATGG
CCATGTCTCTTCCCAACTAGTGCTGGCCTCCTGGCTAGCTGGATTCCTCTGGGTCCTTTGCCCCACTGTC
CTCATGGCCAGCCTGCCTTTCTGTGGCCCCAATGGTATTGACCACTTCTTTCGTGACAGTTGGCCCTTGC
TCAGGCTTTCTTGTGGGGACACCCACCTGCTGAAACTGGCGGCTTTCATGCTCTCTACGTTGGTGTTACT
GGGCTCACTGGCTCTGACCTCAGTTTCCTATGCCTGCATTCTTGCCACTGTTCTCAGGGCCCCTACAGCT
GCTGAGCGAAGGAAAGCGTTTTCCACTTGCGCCTCGCATCTTACAGTGGTGGTCATCATCTATGGCAGTT
CCATCTTTCTCTACATTCGTATGTCAGAGGCTCAGTCCAAACTGCTCAACAAAGGTGCCTCCGTCCTGAG
CTGCATCATCACACCCTTCTTGAACCCATTCATCTTCACTCTCCGCAATGACAAGGTGCAGCAAGCACTG
AGAGAAGCCTTGGGGTGGCCCAGGCTCACTGCTGTGATGAAACAGAGGGTCACAAGTCAAAGGAAATGAT
CTT SEQ ID NO:17

TABLE 62

MNPENWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFLRNFSFLE
LLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRYPAICRPLRYETLMNGHV

SSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRDSWPLLRLSCGDTHLLKLAAFMLSTLVLLGS
LALTSVSYACILATVLRAPTAAERRKAFSTCASHLTVVVIIYGSSIFLYIRMSEAQSKLLNKGASVLSCI
ITPFLNPFIFTLRNDKVQQALREALGWPRLTAVMKQRVTSQRK SEQ ID NO:18

The full amino acid sequence of the NOV9 protein of the invention (SEQ ID NO:18, Query) was found to have 151 of 300 amino acid residues (50%) identical to, and 213 of 300 amino acid residues (71%) similar to, the 313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus musculus (Mouse) (OLFACTORY RECEPTOR C6) (Subject) (Table 63). The full amino acid sequence of the protein of the invention (SEQ ID NO:18, Query) was also found to have 138 of 303 amino acid residues (45%) identical to, and 193 of 303 amino acid residues (63%) similar to, the 311 amino acid residue ptnr: SPTREMBL-ACC:O95007 WUGSC:H_DJ0669B10.3 protein from Homo sapiens (Subject) (Table 64). The full amino acid sequence of the protein of the invention (SEQ ID NO:18, Query) was found to have 151 of 300 amino acid residues (50%) identical to, and 213 of 300 amino acid residues (71%) similar to, the 313 amino acid residue ptnr:SPTREMBL-ACC:Q9Z1V0 protein from Mus musculus (Mouse) (OLFACTORY RECEPTOR C6) (Subject) (Table 65).

TABLE 63

```
Best hits (BLASTP Non-Redundant Composite database):
>ptnr:SPTREMBL-ACC:Q9Z1V0 OLFACTORY RECEPTOR C6 - Mus musculus (Mouse), 313
aa.
Top  Previous Match  Next Match
           Length = 313
 Score = 810 (285.1 bits), Expect = 1.5e-80, P = 1.5e-80
 Identities = 151/300 (50%), Positives = 213/300 (71%)

Query:      5 NWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFLR 64
              N T VT F+LLG   +  +Q L+FLG ++TY +   G  LII ++ +D+RL+  MY+FLR
Sbjct:      3 NSTTVTEFILLGLSDACELQVLIFLGFLLTYFLILLGNFLIIFITLVDRRLYTPMYYFLR 62

Query:     65 NFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRYP 124
              NF+ LE+   +V+ PKML  I+TG  TIS + C +Q++LYFFLGTT+FFLLAVMS DRY
Sbjct:     63 NFAMLEIWFTSVIFPKMLTNIITGHKTISLLGCFLQAFLYFFLGTTEFFLLAVMSFDRYV 122

Query:    125 AICRPLRYETLMNGHVSSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRDSWPL 184
              AIC PLRY T+M+  V  QLV  SW++G L ++ P+ ++    PFCGPN I+HFF D++PL
Sbjct:    123 AICNPLRYATIMSKRVCVQLVFCSWMSGLLLIIVPSSIVFQQPFCGPNIINHFFCDNFPL 182

Query:    185 LRLSCGDTHLLKLAAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFSTCASH 244
              + L C DT L++   F+++   LLG+LA+T+  Y  IL T+L  P+A ER+KAFSTC+SH
Sbjct:    183 MELICADTSLVEFLGFVIANFSLLGTLAVTATCYGHILYTILHIPSAKERKKAFSTCSSH 242

Query:    245 LTVVVIIYGSSIFLYIRMSE-AQSKLLNKGASVLSCIITPFLNPFIFTLRNDKVQQALRE 303
```

```
SEQ ID NO:90
            + VV + YGS IF+Y+R  +  Q +  NK  ++L+ ++TP LNPFI+TLRN +V+Q  RE
Sbjct:   243 IIVVSLFYGSCIFMYVRSGKNGQGEDHNKVVALLNTVVTPTLNPFIYTLRNKQVKQVFRE 302
SEQ ID NO:91
```

TABLE 64

```
Best hit (human sequence) =
>ptnr:SPTREMBL-ACC:O95007 WUGSC:H_DJ0669B10.3 PROTEIN - Homo sapiens (Human),
311 aa.
Top  Previous Match  Next Match
           Length = 311
 Score = 656 (230.9 bits), Expect = 3.1e-64, P = 3.1e-64
 Identities = 138/303 (45%), Positives = 193/303 (63%)
Query:     1 MNPENWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMY 60
             M  EN T+VT F+L+GFP S  ++  +FL  +V YI+T    ++II+L   ++ LH  MY
Sbjct:     1 MELENQTRVTKFILVGFPGSLSMRAAMFLIFLVAYILTVAENVIIILLVLQNRPLHKPMY 60

Query:    61 FFLRNFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSL 120
             FFL N SFLE   ++V VPK+L    + +++ISF  C+IQ Y +  L  T+  LLA M+
Sbjct:    61 FFLANLSFLETWYISVTVPKLLFSFWSVNNSISFTLCMIQLYFFIALMCTECVLLAAMAY 120

Query:   121 DRYPAICRPLRYETLMNGHVSSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRD 180
             DRY AICRPL Y T+M+  +  +L L SW  GF   L    ++ L FCGPN I+HFF D
Sbjct:   121 DRYVAICRPLHYPTIMSHGLCFRLALGSWAIGFGISLAKIYFISCLSFCGPNVINHFFCD 180

Query:   181 SWPLLRLSCGDTHLLKLAAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFST 240
               P+L LSC D  + +L  F+L+ ++ L  L +T +SY CILAT+L  PT   ++KAFST
Sbjct:   181 ISPVLNLSCTDMSITELVDFILALVIFLFPLFITVLSYGCILATILCMPTG--KQKAFST 238

Query:   241 CASHLTVVVIIYGSSIFLYIRMSEAQSKLLNKGASVLSCIITPFLNPFIFTLRNDKVQQA 300
             CASHL VV I Y + IF+Y R     +  +NK  S+   I+TP LNPFI+ LRN +V++A
Sbjct:   239 CASHLVVVTIFYSAIIFMYARPRVIHAFNMNKIISIFYAIVTPSLNPFIYCLRNREVKEA 298

Query:   301 LRE 303  SEQ ID NO:92
             L++
Sbjct:   299 LKK 301  SEQ ID NO:93
```

TABLE 65

```
BLASTX (non redundant) =
>ptnr:SPTREMBL-ACC:Q9Z1V0 OLFACTORY RECEPTOR C6 - Mus musculus (Mouse), 313
aa.
Top  Previous Match  Next Match
           Length = 313
  Plus Strand HSPs:
 Score = 810 (285.1 bits), Expect = 8.9e-80, P = 8.9e-80
 Identities = 151/300 (50%), Positives = 213/300 (71%), Frame = +2

Query:    20 NWTQVTSFVLLGFPSSHLIQFLVFLGLMVTYIVTATGKLLIIVLSWIDQRLHIQMYFFLR 199
             N T VT F+LLG   +  +Q L+FLG ++TY +   G  LII ++ +D+RL+  MY+FLR
Sbjct:     3 NSTTVTEFILLGLSDACELQVLIFLGFLLTYFLILLGNFLIIFITLVDRRLYTPMYYFLR 62

Query:   200 NFSFLELLLVTVVVPKMLVVILTGDHTISFVSCIIQSYLYFFLGTTDFFLLAVMSLDRYP 379
             NF+ LE+   +V+ PKML  I+TG  TIS + C +Q++LYFFLGTT+FFLLAVMS DRY
```

```
Sbjct:    63 NFAMLEIWFTSVIFPKMLTNIITGHKTISLLGCFLQAFLYFFLGTTEFFLLAVMSFDRYV 122

Query:   380 AICRPLRYETLMNGHVSSQLVLASWLAGFLWVLCPTVLMASLPFCGPNGIDHFFRDSWPL 559
             AIC PLRY T+M+  V  QLV  SW++G L ++ P+ ++   PFCGPN I+HFF D++PL
Sbjct:   123 AICNPLRYATIMSKRVCVQLVFCSWMSGLLLIIVPSSIVFQQPFCGPNIINHFFCDNFPL 182

Query:   560 LRLSCGDTHLLKLAAFMLSTLVLLGSLALTSVSYACILATVLRAPTAAERRKAFSTCASH 739
             + L C DT L++   F+++   LLG+LA+T+  Y  IL T+L  P+A ER+KAFSTC+SH
Sbjct:   183 MELICADTSLVEFLGFVIANFSLLGTLAVTATCYGHILYTILHIPSAKERKKAFSTCSSH 242

Query:   740 LTVVVIIYGSSIFLYIRMSE-AQSKLLNKGASVLSCIITPFLNPFIFTLRNDKVQQALRE 916
SEQ ID NO:94
             + VV + YGS IF+Y+R  +  Q +  NK  ++L+ ++TP LNPFI+TLRN +V+Q  RE
Sbjct:   243 IIVVSLFYGSCIFMYVRSGKNGQGEDHNKVVALLNTVVTPTLNPFIYTLRNKQVKQVFRE 302
SEQ ID NO:95
```

NOV9 varies from NOV4 at (bp): 378, 429, 600, 857, 954 and (aa):124, 142, 198, 284, and 316.

Further, possible SNP positions are as follows:

137: C->A(6)

118832467(i), phred 38

118832469(i), phred 45

118832473(i), phred 45

118832477(i), phred 49

118832479(i), phred 38

118832475(i), phred 37

242: G->C(6)

118832467(i), phred 40

118832469(i), phred 33

118832473(i), phred 37

118832477(i), phred 40

118832479(i), phred 33

118832475(i), phred 49

359: G->A(2)
118832469(i), phred 24
118832475(i), phred 36

NOV10

The novel nucleic acid of 1077 nucleotides (designated CuraGen Acc. No. AC019108_H) encoding a novel GPCR-like protein is shown in Table 66. An open reading frame was identified beginning with an atg initiation codon at nucleotides 78-80 and ending with a tag codon at nucleotides 1032-1034. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 66, and the start and stop codons are in bold letters. The encoded protein having 318 amino acid residues is presented using the one-letter code in Table 67.

TABLE 66

TCTCATTTCTCAATTAAGTGCTAAATGCTGGGTGCTCTTTATATCCCCAGAGGGAGAGAGACCAAGGGTGAGAAGAAA
TGTCCAACGCCAGCCTCGTGACAGCATTCATCCTCACAGGCCTTCCCCATGCCCCAGGGCTGGACGCCCTCCTCTTTG
GAATCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACC
TCCACACCCCCATGTACTACTTCCTCACCAACCTGTCCTTCATTGACATGTGGTTCTCCACTGTCAAGGTGCCCAAAA
TGCTGATGACCTTGGTGTCCCCAAGCGGCAGGGCTATCTCCTTCCACAGCTGCGTGGCTCAGCTCTATTTTTTCCACT
TCCTGGGGAGCACCGGATGTTTCCTCTACACAGACACAGTCATGGCCTATGACCGCTATCTGGCTATCTGTCAACCCC
TGCACTACCCAGTGGCCATGAACAGAAGGATGTGTGCAGAAATGGCTGGAATCACCTGGGCCATAGGTGCCACGCACG
CTGCAATCCACACCTCCCTCACCTTCCGCCTGCTCTACTGTGGGCCTTGCCACATTGCCTACTTCTTCTGCGACATAC
CCCCTGTCCTAAAGCTCGCCTGTACAGACACCACCATTAATGAGCTAGTCATGCTTGCCAGCATTGGCATCGTGGCTG
CAGGCTGCCTCATCCTCATCGTTATTTCCTACATCTTCATCGTGGCAGCTGTGTTGCGCATCCGCACAGCCCAGGACC
GGCAGCGGGCCTTCTCCCCCTGCACTGCCCAGCTCACTGGGGTGCTCCTGTACTACGTGCCACCTGTCTGTATCTACC
TGCAGCCTCGCTCCAGTGAGGCAGGAGCTGGGGCCCCTGCTGTCTTCTACACAATCGTAACTCCAATGCTCAACCCAT
TCATTTACACTTTGTGGAACAAGGAGGTGAAGCATGCTCTGCAAAGGCTTTTGTGCAGCAGCTTCCGAGAGTCTACAG
CAGGCAGCCCACCCCCATAGTCTGTGCTATCAAAACTCACAATTTGCCTGCCAGGAAAGCNGC SEQ ID NO:19

TABLE 67

MSNASLVTAFILTGLPHAPGLDALLFGIFLVVYVLTVLGNLLILLVIRVDSHLHTPMYYFLTNLSFIDMWFSTVKVPK
MLMTLVSPSGRAISFHSCVAQLYFFHFLGSTGCFLYTDTVMAYDRYLAICQPLHYPVAMNRRMCAEMAGITWAIGATH
AAIHTSLTFRLLYCGPCHIAYFFCDIPPVLKLACTDTTINELVMLASIGIVAAGCLILIVISYIFIVAAVLRIRTAQD
RQRAFSPCTAQLTGVLLYYVPPVCIYLQPRSSEAGAGAPAVFYTIVTPMLNPFIYTLWNKEVKHALQRLLCSSFREST
AGSPPP SEQ ID NO:20

In a search of sequence databases, it was found, for example, that NOV10 (Query) has 607 of 957 bases (63%) identical to a Rat species GPCR mRNA (GENBANK-ID: U50949) (Subject) (Table 68). NOV10 (Query) was also found to have 174 of 299 amino acid residues (58%) identical to, and 217

of 299 residues (72%) positive with, the 318 amino acid residue protein from Rat species (ptnr:SPTREMBL-ACC:Q62944) (Subject) (Table 69).

TABLE 68

```
gb:GENBANK-ID:RNU50949|acc:U50949 Rattus norvegicus taste bud receptor protein
TB 641 (TB 641) gene, complete cds - Rattus norvegicus, 1278 bp. Length = 1278
  Minus Strand HSPs:
 Score = 1347 (202.1 bits), Expect = 6.8e-55, P = 6.8e-55
 Identities = 607/957 (63%), Positives = 607/957 (63%), Strand = Minus / Plus

Query:  1040 TTTATATCCCCAGAGGGAGAGA-GACCAAGGGTGAGAAGAAATGTCCAACGCCAGCCTCG 982
             TT AT T   CAGA   AGA A GA  AAG    AGAA A A  TC    G CA   T G
Sbjct:   115 TTCATCTTAACAGATTAAGACATGAG-AAGAAACAGAA-ACACATCACTGGACACTGTGG 172

Query:   981 TGACAGCATTCATCCTCACAGGCCTTCCC-CATGCCCCAGGGCTG-GACGCCCTCCTCTT 924
             TGACAG  TTC T CTC C GG CTT  C CAT CCCCA   CT  GA  C  TCCTCTT
Sbjct:   173 TGACAGATTTCCTTCTC-CTGGGCTTGGCTCATCCCCCAAATCTAAGAA-CGTTCCTCTT 230

Query:   923 TGGAATCTTCCTGGTGGTTTACGTGCTCACTGTGCTGGGGAACCTCCTCATCCTGCTGGT 864
                  TCTTCCT  T  TTTAC T CT AC   G TGGGGAACCT CTCATCCTGCT
Sbjct:   231 CCTGGTCTTCCTCCTCATTTACATCCTGACACAGTTGGGGAACCTGCTCATCCTGCTCAC 290

Query:   863 GATCAGGGTGGATTCTCACCTCCACACCC-CCATGTACTACTTCCTCACCAACCTGTCCT 805
               T  GGG  GA  C  A CT CA  CCC CC   T  TAC T CT      C TG  CT
Sbjct:   291 AGTGTGGGCTGACCCCAAGCTGCATGCCCGCCCCATG-TACATTCTGCTGGGCGTGCTCT 349

Query:   804 TCAT--TGACATGTGGTTCTCCACTGTCAAGGTG-CCCAAAATGCTGATGACCTTGGTGT 748
              C T  TG  A  TG   CTC  CT       TG CCC   A      T A   T  T
Sbjct:   350 CCTTCCTGG-ACATGTGGCTCTCCTCAGTCATTGTCCCTCGAATTATTTTAAACT--TCA 406

Query:   747 CCCCAAGCGGCAGGGCTATCTCCTTCCACAGCTGCGTGGCTCAGCTCTATTTTTTCCACT 688
             C CC   C  CA GGCTATC C TT     GCTG GT GCTCA CTCTATTTTTTCCACT
Sbjct:   407 CTCCTGCCAACAAGGCTATCGCATTTGGTGGCTGTGTAGCTCAACTCTATTTTTTCCACT 466

Query:   687 TCCTGGGGAGCACCGGATGTTTCCTCTACACAGACACAGTCATGGCCTATGACCGCTATC 628
             TCCTGGG AGCACC   TG TTCCTCTA AC   C    T ATGGCCTATGAC G TA C
Sbjct:   467 TCCTGGGCAGCACCCAGTGCTTCCTCTATAC---CT---TGATGGCCTATGACAGGTACC 520

Query:   627 TGGCTATCTGTCAACCCCTGCACTACCCAGTGGCCATGAACAGAAGGATGTGTGCAGAAA 568
             TGGC AT TGTCA CC CT C CTACCC GTG  CATGAA  G A G T TG  CA
Sbjct:   521 TGGCAATATGTCAGCCTCTTCGCTACCCTGTGCTCATGAATGGGAAGTTATGCACAATCC 580

Query:   567 TGGCTGGAATCACCT-GGGCCATAGGTGCCACGCACGCTGCAATCCACACCTCCCTCACC 509
             TGG TGG    A CT GGG     GG  CCA  CA G   C AT CA  CC C CT ACC
Sbjct:   581 TGG-TGGCTGGAGCTTGGGTGGCTGGCTCCATCCATGGGTCTATTCAAGCCACTCTGACC 639

Query:   508 TTCCGCCTGCTCTACTGTGGGCCTTGCCACATTGCCTACTTCTTCTGCGACATACCCCCT 449
             TTCCG  TGC CTACTGTGGGCCT    A  T G  TACTTCTTCTG GACAT CC  C
Sbjct:   640 TTCCGATTGCCCTACTGTGGGCCTAAGGAAGTGGATTACTTCTTCTGTGACATTCCTGCA 699

Query:   448 GTCCTAAAGCTCGCCTGTACAGACACCACCATTAATGAGCTAGTCATGCTTGCCAGCATT 389
             GT CT A  CT GCCTGT C GA AC  C AT AATGA CT GT A   TTG    CATT
Sbjct:   700 GTGCTGAGACTGGCCTGTGCTGATACAGCAATCAATGAACTGGTGACCTTTGTGGACATT 759

Query:   388 GGCATCGTGGCTGCAGGCTGCCTCATCCTCATCGTTATTTCCTACATCTTCATCGTGGCA 329
```

```
              GG  T GTGGCTGC  G TGC TC T CT AT  T  T TCCTAC  C  CAT GT  CA
Sbjct:   760 GGGGTAGTGGCTGCCAGTTGCTTCCTGCTGATTCTGCTCTCCTACGCCAACATAGTT-CA 818

Query:   328 -GCTGTGTTGCGCATCCGCACAGCCCAGGACCGGCAG-CGGGCCTTCTCCCCCTGCA-CT 272
              GC  T  TG   AT CGCAC GC  A G C GG AG CG GCCTTCTCC CCTG   CT
Sbjct:   819 TGCCATCCTGAAGATACGCACTGCAGATGGCAGG-AGACGTGCCTTCTCCACCTGTGGCT 877

Query:   271 GCCCAGCTCACTGGGGTGCTCC-TGTACTACGTGCCACCTGTCTGT--ATCTACCTGCAG 215
              CCCA CTCACTG GGT C C  T TACTA GT CC CCTGT T T  ATCTACCT C G
Sbjct:   878 -CCCATCTCACTGTGGT-CACAGTCTACTATGT-CC-CCTGTATTTTCATCTACCTTCGG 933

Query:   214 CCTCGCTCCAGTGAGG-CAGGAGCTGGGGCCCCTGCTGTCTTCTACACAATCGTAACTCC 156
              C  G TCCA  GAG  C    G  GG GC   TGCTGT TT TACAC  T GT ACTCC
Sbjct:   934 GCAGGTTCCAA-GAGTTCCTTTGACGGAGCAGTTGCTGTATTTTACACTGTTGTCACTCC 992

Query:   155 AATGCTCAACCCATTCATTTACACTTTGTGGAACAAGGAGGTGAAGCATGCTCTGCAAAG 96
              A T CT AA CC  TCAT TACACT TG GGAAC AGGA GTGAA   TGC CTG A AG
Sbjct:   993 ATTACTGAATCCCCTCATCTACACTCTGAGGAACCAGGAAGTGAATTCTGCCCTGAAGAG 1052

Query:    95 GCTTTTGTGCAG 84  SEQ ID NO:96
             GCT   G GCAG
Sbjct:  1053 GCTAA-GAGCAG 1063  SEQ ID NO:97
```

TABLE 69

ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - RATTUS NORVEGICUS (RAT), 318 aa. Length = 318

Minus Strand HSPs:

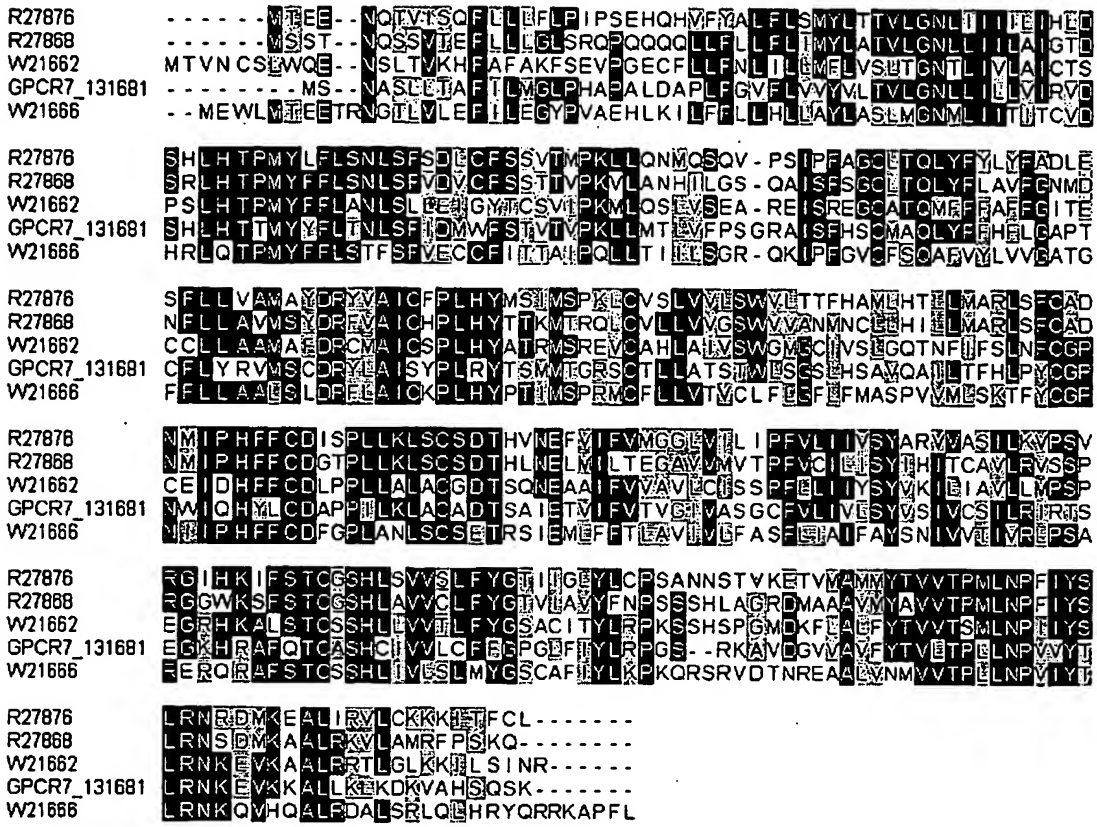
Score = 896 (315.4 bits), Expect = 5.2e-89, P = 5.2e-89

Identities = 174/299 (58%), Positives = 217/299 (72%), Frame = -3

```
Query:   988 SLVTAFILTGLPHAPGLDALLFGIFLVVYVLTVLGNLLILLVIRVDSHLHT-PMYYFLTN 812
              ++VT F+L GL H P L   LF +FL++Y+LT LGNLLILL +  D  LH  PMY  L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   811 LSFIDMWFSTVKVPKMLMTLVSPSGRAISFHSCVAQLYFFHFLGSTGCFLYTDTVMAYDR 632
              LSF+DMW S+V VP++++   +P+ +AI+F  CVAQLYFFHFLGST CFLYT  +MAYDR
Sbjct:    71 LSFLDMWLSSVIVPRIILNF-TPANKAIAFGGCVAQLYFFHFLGSTQCFLYT--LMAYDR 127

Query:   631 YLAICQPLHYPVAMNRRMCAEMAGITWAIGATHAAIHTSLTFRLLYCGPCHIAYFFCDIP 452
              YLAICQPL YPV MN ++C  +    W  G+ H +I  +LTFRL YCGP  + YFFCDIP
Sbjct:   128 YLAICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIP 187

Query:   451 PVLKLACTDTTINELVMLASIGIVAAGCLILIVISYIFIVAAVLRIRTAQDRQRAFSPCT 272
              VL+LAC DT INELV    IG+VAA C +LI++SY  IV A+L+IRTA  R+RAFS C
Sbjct:   188 AVLRLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCG 247

Query:   271 AQLTGVLLYYVPPVCIYLQPRSSEAGAGAPAVFYTIVTPMLNPFIYTLWNKEVKHALQRL 92
SEQ ID NO:98
              + LT V +YYVP + IYL+  S  +  GA AVFYT+VTP+LNP IYTL N+EV  AL+RL
Sbjct:   248 SHLTVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRL 307
SEQ ID NO:99
```

A multiple sequence alignment is given in Table 70, with the protein of the invention being shown on line 4, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV10 protein is a novel member of the OR family. In the following table, GPCR7_131681 is the full-length SEQ ID NO:20. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 70



Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (Table 71).

TABLE 71

```
plasma membrane --- Certainty=0.4600(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.2408(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.1000(Affirmative) < succ>
endoplasmic reticulum (lumen) --- Certainty=0.1000(Affirmative) < succ>
```

SignalP analysis of the protein of invention indicated that it has got secretory signal peptide (Table 72). The first 70 amino acids of NOV 10 (SEQ ID NO:20) were used for signal peptide prediction.

TABLE 72

```
< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C     40       0.804   0.37   YES
  max. Y     52       0.649   0.34   YES
  max. S     33       0.982   0.88   YES
  mean S      1-51    0.784   0.48   YES
# Most likely cleavage site between pos. 51 and 52: VDS-HL
```

NOV11

The novel nucleic acid of 960 nucleotides (designated CuraGen Acc. No. nh0413n10_A_da2; NOV11; SEQ ID NO:21) encoding a novel OR-like protein (SEQ ID NO:22) is shown in Table 73. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1, 2 and 3 and ending with a TGA codon at nucleotides 943, 944 and 945. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 73, and the start and stop codons are in bold letters. The encoded protein having 314 amino acid residues is presented using the one-letter code in Table 74.

TABLE 73

```
ATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGACCC
AGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCATCAT
TCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTCTTAGAAATATCTCCTTTGCAGAT
CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTT
TTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCAGT
GATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTG
TGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTT
TCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCT
GGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCT
ATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGA
GACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATT
CACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCG
GTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAAC
TAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTCTGAGTCTGACT  SEQ ID NO:21
```

TABLE 74

```
MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISFAD
LCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIMTQRV
CLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGVVILLAP
ISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTA
VTPMLNPIIYSLRNKDVKGALRKLVGRKCFSHRQ  SEQ ID NO:22
```

In a search of sequence databases, it was found, for example, that NOV11 (Query) has 526 of 648 bases (81 %) identical to a *Mus musculus* olfactory receptor gene mRNA

(GENBANK-ID: AF073987|acc:AF073987) (Subject) (Table 75). NOV11 (Query) was also found to have 167 of 307amino acid residues (54%) identical to, and 218 of 307 residues (71 %) positive with, the 317 amino acid residue OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 protein from Canis familiaris (Dog) (ptnr:SWISSPROT-ACC: Q95156) (Subject) (Table 76). Furthermore, NOV11 differs at positions 196 and 631, and the amino acid sequence differs at residues 66 & 211 from that given in NOV12 below.

TABLE 75

```
>gb:GENBANK-ID:AF073987|acc:AF073987 Mus musculus domesticus clone OR912-47M9
        olfactory receptor gene, partial cds - Mus musculus domesticus, 649
bp.

Top  Previous Match  Next Match

            Length = 649

  Plus Strand HSPs:

 Score = 2126 (319.0 bits), Expect = 6.1e-90, P = 6.1e-90
 Identities = 526/648 (81%), Positives = 526/648 (81%), Strand = Plus / Plus

Query:   201 CTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGT 260
             CTTTGCAGATCTCTG TT TCTACTA CA  GT CC CA GTG T GT CACTTC TGGT
Sbjct:     1 CTTTGCAGATCTCTGCTTTTCTACTACCACAGTGCCCCAGGTGCTTGTCCACTTCCTGGT 60

Query:   261 AAAGAGGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGT 320
              AAGAGGAA ACCATTTCTTTT  TGG TGT   ACACAGATA T GT TT CTTCTGGT
Sbjct:    61 GAAGAGGAAGACCATTTCTTTTGCTGGATGTTCTACACAGATAGTGGTGTTGCTTCTGGT 120

Query:   321 TGGGTGTACAGAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTG 380
              GG TG ACAGAGTGTGC CTGCTGGCAGTGATGTCCTATGACCG TATGTGGCTGTCTG
Sbjct:   121 CGGATGCACAGAGTGTGCACTGCTGGCAGTGATGTCCTATGACCGATATGTGGCTGTCTG 180

Query:   381 CAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAG 440
             CAA CC CTG ACTACTC ACCATCATGACACA  GG T TGT T   GCTG C T CAG
Sbjct:   181 CAAACCTCTGCACTACTCCACCATCATGACACACTGGCTATGTGTTCAGCTGGC-TGCAG 239

Query:   441 G-TCCTGGGCCAGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTC 499
             G TCCTGGGCCAGTGG GCACT GTGTC  T GT GATACCA  TT AC TT C TCTTC
Sbjct:   240 GGTCCTGGGCCAGTGGTGCACTTGTGTCCCTGGTGGATACCACATTCACATTACGTCTTC 299

Query:   500 CCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGC 559
             C TA  G GGA A AAT T AT AA CACT  TT TGTGAACCTCCTGCCCTCCTGAAGC
Sbjct:   300 CTTATCGAGGAAACAATGTCATTAACCACTTTTTCTGTGAACCTCCTGCCCTCCTGAAGC 359

Query:   560 TGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCC 619
             TGGC TC   AGA AC TACAGCACAGA ATGGC ATCTTT CAATGGG GTGGTAATCC
Sbjct:   360 TGGCATCGGCAGATACATACAGCACAGAGATGGCGATCTTTGCAATGGGTGTGGTAATCC 419

Query:   620 TCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCC 679
```

```
              TCCT GC CCT TCTCCCT AT CT    TC TA TGGAA AT ATCTCCACTGT ATCC
Sbjct:   420 TCCTAGCACCTGTCTCCCTCATCCTCACCTCCTACTGGAACATCATCTCCACTGTAATCC 479

Query:   680 AGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTG 739
             AGATGCAGTCTGGGGAAGG AG CTCAAGG  TT TCCACCTGTGGCTCCCA CT ATTG
Sbjct:   480 AGATGCAGTCTGGGGAAGGAAGGCTCAAGGTCTTCTCCACCTGTGGCTCCCACCTCATTG 539

Query:   740 TTGTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTA 799
             TTGTTGT CTCTTCTA GG TCAG AATATT  CCTACATG G CC AACTC AAGA  A
Sbjct:   540 TTGTTGTTCTCTTCTACGGCTCAGCAATATTTGCCTACATGAGGCCCAACTCTAAGATAA 599

Query:   800 CAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCC 848  SEQ ID
NO:104
               AA GAA  GGATAAAATGAT TC GTGTTCTAT CAGC GTGAC CC
Sbjct:   600 TGAATGAAAAGGATAAAATGATTTCGGTGTTCTATTCAGCAGTGACCCC 648  SEQ ID
NO:105
```

TABLE 76

```
>ptnr:SWISSPROT-ACC:Q95156 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Canis
        familiaris (Dog), 317 aa.


Top  Previous Match  Next Match

           Length = 317

  Plus Strand HSPs:

 Score = 844 (297.1 bits), Expect = 2.2e-83, P = 2.2e-83
 Identities = 167/307 (54%), Positives = 218/307 (71%), Frame = +1

Query:     1 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 180
             MG  NQT+V +F+ LGLS D  T++ LF+LFLI Y++TVLGN LII+LI LDSRLHTPMY
Sbjct:     1 MGTGNQTWVREFVLLGLSSDWDTEVSLFVLFLITYMVTVLGNFLIILLIRLDSRLHTPMY 60

Query:   181 FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 360
             FFL N+S  D+ ++TSI+PQ+L H L   K I F  C  Q+   L +G  E  LLAVM+Y
Sbjct:    61 FFLTNLSLVDVSYATSIIPQMLAHLLAAHKAIPFVSCAAQLFFSLGLGGIEFVLLAVMAY 120

Query:   361 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 540
             DRYVAVC PL YS IM   +C  L+  SW SG++ SL+ T  TF LP      I+H  CE
Sbjct:   121 DRYVAVCDPLRYSVIMHGGLCTRLAITSWVSGSMNSLMQTVITFQLPMCTNKYIDHISCE 180

Query:   541 PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST 720
               A+++LA +DT S E+AI    +V+L+ P  L+L SY  IIST++++QS EGR KAF T
Sbjct:   181 LLAVVRLACVDTSSNEIAIMVSSIVLLMTPFCLVLLSYIQIISTILKIQSTEGRKKAFHT 240

Query:   721 CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA 900
             C SHL VVVL YG  IFTY++P S  +   +K+IS+FY+ +TPMLNP+IYS+RNK+VKGA
Sbjct:   241 CASHLTVVVLCYGMAIFTYIQPRSSPSVLQEKLISLFYSVLTPMLNPMIYSVRNKEVKGA 300

Query:   901 LRKLVGR 921  SEQ ID NO:106
              +KL+G+
Sbjct:   301 WQKLLGQ 307  SEQ ID NO:107
```

A multiple sequence alignment is given in Table 77, with the protein of the invention being shown on line 1, in a ClustalW analysis comparing the protein of the invention with 2 related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. The top sequence is NOV11; the middle sequence is Q95156: ptnr:SWISSPROT-ACC:Q95156 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Canis familiaris (Dog), (317 aa); and the bottom sequence is Q13607: ptnr:SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens(Human), (317 aa). Based on its relatedness to members of the GPCR family, the NOV11 protein is a novel member of the OR family. In the following table, nh0413n10_A_da2 is the full-length SEQ ID NO:22. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 77

Tissue expression

The OR disclosed in this invention is expressed in at least one of the following tissues: adrenal gland, bone marrow, brain – amygdala, brain – cerebellum, brain – hippocampus, brain - substantia nigra, brain – thalamus, brain – whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma – Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus.

Localization / Sorting

The SignalP, Psort and or hydropathy profile indicate that this sequence has a signal peptide and is likely to be localized at the membrane (see Table 79A, B and C).

TABLE 79 A

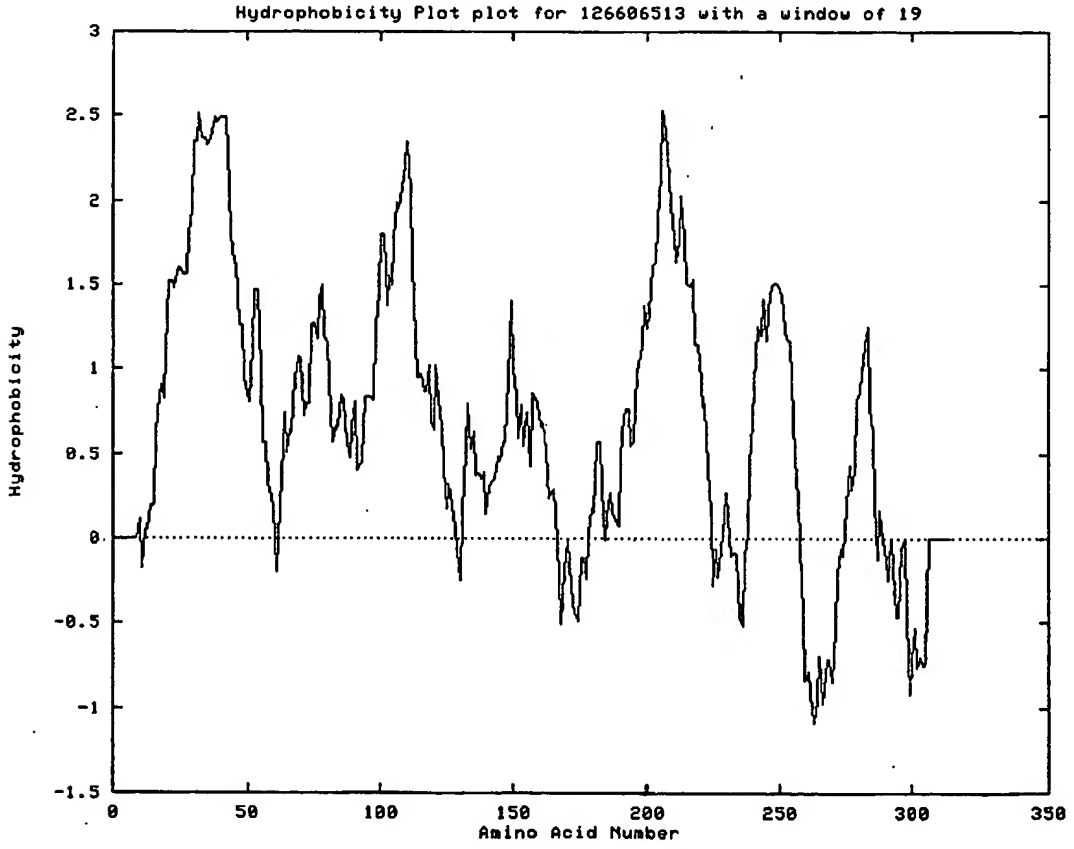


TABLE 79B

```
                plasma membrane --- Certainty=0.6000(Affirmative) < succ>
                    Golgi body --- Certainty=0.4000(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3000(Affirmative) < succ>
          microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
```

TABLE 79C

Signal P

The first 70 amino acids of 126606513 (314 aa) were used for signal peptide prediction

```
< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C     42       1.000   0.37   YES
  max. Y     42       0.629   0.34   YES
  max. S     35       0.979   0.88   YES
  mean S      1-41    0.593   0.48   YES
# Most likely cleavage site between pos. 41 and 42: VLG-NQ
```

NOV12

NOV12 is 994 nucleotides (designated CuraGen Acc. No. AC0170103_B; SEQ ID NO: 23) and encodes a novel Olfactory Receptor -like protein (SEQ ID NO:24) is shown in Table 80. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 31-33 and ending with a TGA codon at nucleotides 973-75. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 80, and the start and stop codons are in bold letters. The encoded protein having 314 amino acid residues is presented using the one-letter code in Table 81.

TABLE 80

TGCCAAACAGGTAAACAGGCAAAAATATCA**ATG**GGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGT
CTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAAC
CAGCTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATCTCTCCTTT
GCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTT
TATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCAGTGATGTCCTAT
GACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTC
AGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAAT
ATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCC
ATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACT
GTTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTC
TTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTG

TTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGG
AAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA SEQ ID NO:23

TABLE 81

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISFADLCFSTSIV
PQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIMTQRVCLWLSFRSW
ASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIIS
TVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNK
DVKGALRKLVGRKCFSHRQ SEQ ID NO:24

In a search of sequence databases, it was found, for example, that NOV12 (Query) has 617 of 930 bases (66 %) identical to a *Homo sapiens* species Olfactory Receptor clone (GENBANK-ID: AC004853) (Subject) (Table 82). The full amino acid sequence of the protein (SEQ ID NO:24; Query) of the invention was found to have 168 of 307 amino acid residues (54 %) identical to, and 218 of 307 residues (71%) positive with, the 317 amino acid residue proteins from *Canis familiaris* (ptnr: SPTREMBL-ACC: Q95156) (Subject) (Table 83).

TABLE 82

```
>gb:GENBANK-ID:AC004853|acc:AC004853 Homo sapiens PAC clone DJ0669B10 from
7q33-q35,
complete sequence - Homo sapiens, 129355 bp.

Score = 1625 (243.8 bits), Expect = 3.5e-66, P = 3.5e-66
Identities = 617/930 (66%), Positives = 617/930 (66%), Strand = Plus / Plus

Query:    22 AAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTT 81
             AA AT T AATGGGA  AGA AACCA AC T  GTG    A TTTAT  TCCT GG CT
Sbjct: 41891 AATATTTTAATGGGAACAGATAACCAGACTTGGGTGAGTGAATTTATTCTCCTCGGCCTG
41950

Query:    82 TCACAG-GACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCT 140
             TC CAG GACT G A AC C G TC   CT TTT TCCT TTC T  TCAT TAT TG T
Sbjct: 41951 TC-CAGTGACTGGGACACTCGGGTCTCCCTGTTTGTCCTGTTCTTGGTCATGTATGTGGT
42009

Query:   141 GACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACAC 200
             GACCGTGCT GG AAC   CTCAT  TC TTCT ATC   CTGGA    CG CT CACAC
Sbjct: 42010 GACCGTGCTGGGGAACTGTCTCATTGTCCTTCTGATCAGACTGGACAGCCGACTCCACAC
42069

Query:   201 TCCCATGTATTTTTTTCTTAGAAATCTCTCCTTTGCAGATCTCTGTTTCTCTACTAGCAT 260
             TCCCATGTATTT TTTCT A  AA CTCTCC TTG  GAT TCT  T   C AC AG  T
Sbjct: 42070 TCCCATGTATTTCTTTCTCACCAACCTCTCCCTTGTCGATGTCTCCTATGCCACAAGTGT
42129

Query:   261 TGTCCCTCAAGTGTTGGTTCACTT-CTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGT 319
              GTCCCTCA  TG TGG  CA TT CTTG   AA A  AAA CCAT  C TT  A  G T
```

```
Sbjct: 42130 AGTCCCTCAGCTGCTGGCACATTTTCTTGCAGAACAT-AAAGCCATCCCATTCCAGAGCT
42188

Query:   320 GTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGT-ACAGAGTGTGCGCTGCTGGCA 378
             GT    C CAG TA TT TCT  CT     TTGGGTG  A  GAGT TG  CT CTGGC
Sbjct: 42189 GTGCAGCCCAGTTATTTTTCTCCCTGGCC-TTGGGTGGGATTGAGTTTGTTCTCCTGGCG
42247

Query:   379 GTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATG 438
             GTGATG CCTATGACCG TATGTGGCTGT TG  A  CCCTG   TACTC  CCATCATG
Sbjct: 42248 GTGATGGCCTATGACCGCTATGTGGCTGTGTGTGATGCCCTGCGATACTCGGCCATCATG
42307

Query:   439 ACACAACGGG-TGTGTCTCT-GGCTGTCCTTCAGGTCCTGGGCCAGTGGGG-CACTAGTG 495
              CA    GGG TGTGT  CT GG TG CC TCA  TCCTGGG CAGTGG   CA  AG
Sbjct: 42308 -CATGGAGGGCTGTGTG-CTAGGTTGGCCATCACATCCTGGGTCAGTGGCTTCATCAGC-
42364

Query:   496 TCTTTAGTAGATACCAGCTTT-ACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAA 554
             TCT   GT  A AC  GCT T AC TT CA CT CCC   TG  GA A AA  T AT  A
Sbjct: 42365 TCTCCTGTGCAGACT-GCTATCACCTTTCAGCTGCCCATGTGCAGAAACAAGTTTATTGA
42423

Query:   555 TCAC-TACTTTTGTGAACCTCCT-GCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGC 612
             TCAC TA T  TGTGAAC TCCT GC  T  T A GCTGGCTT   T GACAC T C  C
Sbjct: 42424 TCACATA-TCCTGTGAAC-TCCTAGCTGTGGTCAGGCTGGCTTGTGTGGACACCTCCTCC
42481

Query:   613 ACAGAAATGGCCATCTTT-TCAATGGGCGTGGTAATCCTCCTGGC-CCCTGTCTCCCTGA 670
             A  GA  T  CCATC T  T   T G C T GT  T CT  TG C CCCT TCT CCTG
Sbjct: 42482 AATGAGGTCACCATCATGGTGTCTAG-CATTGTTCTTCTGATGACACCCT-TCTGCCTGG
42539

Query:   671 TTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGA 730
             TTCTT   TC TA     A AT ATCTCCAC  T  T  AGAT CAGTC  G GAAGG A
Sbjct: 42540 TTCTTTTGTCCTACATCCAGATCATCTCCACCATCCTAAAGATCCAGTCCAGAGAAGGAA
42599

Query:   731 GACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGT 790
             GA   AA GCTTT   CAC TGTG CTC CA CT A  GT GTTG CCT T CTATGG T
Sbjct: 42600 GAAAGAAAGCTTTCCACACGTGTGCCTCTCACCTCACAGTGGTTGCCCTGTGCTATGG-T
42658

Query:   791 CAGGA-ATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATG 849
               GG  AT TTCAC TACAT C  CC  ACTCCA   C  C       C GGA AA  TG
Sbjct: 42659 GTGGCCATTTTCACTTACATCCAGCCCCACTCCAGTCCCTCTGTCCTTCAGGAGAAGTTG
42718

Query:   850 ATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGG 909
              T TCTGT TT TAT C     T AC CCAATG TGAACCCCAT ATTTA AGC T AGG
Sbjct: 42719 TTCTCTGTCTTTTATGCCATTTTAACACCAATGCTGAACCCCATGATTTACAGCCTAAGG
42778

Query:   910 AACAAAGATGTCAAAGGGGC-TCTCAGGAAACTAGTTGGGAGA 951  SEQ ID NO:110
             AA AAAGA GT AA GGGGC T  CAG AAACTA T  GGA A
Sbjct: 42779 AATAAAGAGGTGAAGGGGGCCTGGCAG-AAACTATTATGGAAA 42820 SEQ ID NO:111
```

TABLE 83

```
ptnr:SWISSPROT-ACC:Q95156 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Canis
familiaris
 (Dog), 317 aa.

Score = 845 (297.5 bits), Expect = 8.5e-84, P = 8.5e-84
Identities = 168/307 (54%), Positives = 218/307 (71%), Frame = +2

Query:    86 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 265
             MG  NQT+V +F+ LGLS D  T++ LF+LFLI Y++TVLGN LII+LI LDSRLHTPMY
Sbjct:     1 MGTGNQTWVREFVLLGLSSDWDTEVSLFVLFLITYMVTVLGNFLIILLIRLDSRLHTPMY 60

Query:   266 FFLRNLSFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 445
             FFL NLS  D+ ++TSI+PQ+L H L   K I F  C  Q+   L +G  E  LLAVM+Y
Sbjct:    61 FFLTNLSLVDVSYATSIIPQMLAHLLAAHKAIPFVSCAAQLFFSLGLGGIEFVLLAVMAY 120

Query:   446 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 625
             DRYVAVC PL YS IM   +C  L+  SW SG++ SL+ T  TF LP      I+H  CE
Sbjct:   121 DRYVAVCDPLRYSVIMHGGLCTRLAITSWVSGSMNSLMQTVITFQLPMCTNKYIDHISCE 180

Query:   626 PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST 805
               A+++LA +DT S E+AI    +V+L+ P  L+L SY  IIST++++QS EGR KAF T
Sbjct:   181 LLAVVRLACVDTSSNEIAIMVSSIVLLMTPFCLVLLSYIQIISTILKIQSTEGRKKAFHT 240

Query:   806 CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA 985
             C SHL VVVL YG  IFTY++P S  +   +K+IS+FY+ +TPMLNP+IYS+RNK+VKGA
Sbjct:   241 CASHLTVVVLCYGMAIFTYIQPRSSPSVLQEKLISLFYSVLTPMLNPMIYSVRNKEVKGA 300

Query:   986 LRKLVGR 1006 (SEQ ID NO:112)
               +KL+G+
Sbjct:   301 WQKLLGQ 307  (SEQ ID NO: 113)
```

NOV13

In the present invention, the target sequence identified above as NOV12 (Accession Number AC0170103_B), was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such suitable sequences were then employed as the forward and reverse primers in a PCR amplification based on a library containing a wide range of cDNA species. The resulting amplicon was gel purified, cloned and sequenced to high redundancy to provide the sequence reported below, which is designated NOV13 (Accession Number nh0413n10_A1). NOV13 is 994 nucleotides (SEQ ID NO: 25) (Table

85) encoding a novel GPCR-like protein (SEQ ID NO:26) (Table 86). An open reading frame was identified beginning with an ATG initiation codon at nucleotide 31 and ending with a TAG codon at nucleotide 972. The encoded protein has 314 amino acid residues. There is 1 amino acid difference for nh0413n10_A1 with respect to NOV12 (position 66, Leu becomes Ile).

TABLE 85

TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCT
TTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGC
TCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGAT
CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTG
TATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATG
TGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCC
AGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTA
CTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCG
TGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCT
GGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATT
CACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAA
TGTTGAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTC
TCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA SEQ ID NO:25

TABLE 86

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNISFADLCFSTSIV
PQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIMTQRVCLWLSFRSW
ASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIIS
TVIQMQSGEGRLKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNK
DVKGALRKLVGRKCFSHRQ SEQ ID NO:26

Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (certainity =0.6000). The polypeptide seems to have no cleavable N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 41 and 42.

In a search of sequence databases, it was found, for example, that NOV13 (SEQ ID NO:26; Query) was found to have 194 of 216 amino acid residues (89%) identical to Mus musculus domesticus (western European house mouse) (ACC:AAD43436 OLFACTORY RECEPTOR) (Subject) and 194 of 216 residues (89%) positive with, the 216 amino acid residue protein (Table 87). The full amino acid sequence of NOV13 (Query) was found to have 165 of 305 amino acid residues (54%) identical to ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens (Subject) and 217 of 305 residues (71%) positive with, the 317 amino acid residue protein (Table 88).

TABLE 87

ptnr:TREMBLNEW-ACC:AAD43436 OLFACTORY RECEPTOR - Mus musculus domesticus (western European house mouse), 216 aa (fragment).

Top Previous Match Next Match
Length = 216

```
 Score = 941 (331.2 bits), Expect = 1.9e-94, P = 1.9e-94
 Identities = 182/216 (84%), Positives = 194/216 (89%)

Query:    68 FADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVC 127
            FADLCFST+ VPQVLVHFLVKRKTISF GC TQI+V LLVGCTECALLAVMSYDRYVAVC
Sbjct:     1 FADLCFSTTTVPQVLVHFLVKRKTISFAGCSTQIVVLLLVGCTECALLAVMSYDRYVAVC 60

Query:   128 KPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKL 187
            KPL+YSTIMT  +C+ L+  SWASGALVSLVDT+FT  LPY G N+INH+FCEPPALLKL
Sbjct:    61 KPLHYSTIMTHWLCVQLAAGSWASGALVSLVDTTFTLRLPYRGNNVINHFFCEPPALLKL 120

Query:   188 ASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIV 247
            AS DTYSTEMAIF+MGVVILLAP+SLIL SYWNIISTVIQMQSGEGRLK FSTCGSHLIV
Sbjct:   121 ASADTYSTEMAIFAMGVVILLAPVSLILTSYWNIISTVIQMQSGEGRLKVFSTCGSHLIV 180

Query:   248 VVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTP 283 SEQ ID NO:112
            VVLFYGS IF YMRPNSK   E DKMISVFY+AVTP
Sbjct:   181 VVLFYGSAIFAYMRPNSKIMNEKDKMISVFYSAVTP 216  SEQ ID NO:113
```

TABLE 88

SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens (Human), 317 aa.

Length = 317

```
 Score = 840 (295.7 bits), Expect = 9.8e-84, P = 9.8e-84
 Identities = 165/305 (54%), Positives = 217/305 (71%)

Query:     1 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 60
             MG +NQT+VS+FI LGLS D  T++ LF+LFL++Y++TVLGN LI++LI LDSRLHTPMY
Sbjct:     1 MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60

Query:    61 FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 120
             FFL N+S  D+ ++TS+VPQ+L HFL + K I F  C  Q+   L +G  E  LLAVM+Y
Sbjct:    61 FFLTNLSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120

Query:   121 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 180
             DRYVAVC  L YS IM   +C  L+  SW SG + S V T+ TF LP      I+H  CE
Sbjct:   121 DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180

Query:   181 PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST 240
               A+++LA +DT S E+ I    +V+L+ P+ L+L SY  IIST++++QS EGR KAF T
Sbjct:   181 LLAVVRLACVDTSSNEVTIMVSSIVLLMTPLCLVLLSYIQIISTILKIQSREGRKKAFHT 240

Query:   241 CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA 300
             C SHL VV L YG  IFTY++P+S  +   +K+ SVFY  +TPMLNP+IYSLRNK+VKGA
Sbjct:   241 CASHLTVVALCYGVAIFTYIQPHSSPSVLQEKLFSVFYAILTPMLNPMIYSLRNKEVKGA 300

Query:   301 LRKLV 305  SEQ ID NO:114
              +KL+
Sbjct:   301 WQKLL 305  SEQ ID NO:115
```

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs (Table 84).

TABLE 89

Possible SNPs found:
200: T->C(2)
127182648(i), phred 36
128123424(i), phred 36
221: A->G(2)
127182619(i), phred 40
128123383(i), phred 29
301: A->G(2)
127182619(i), phred 38
128123383(i), phred 33
330: G->A(2)
127182648(i), phred 26
128123424(i), phred 26

NOV14

In the present invention, the target sequence identified as NOV5, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their

identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. NOV14 is 1012 nucleotides (Table 91) encoding a novel GPCR-like protein (Table 92). An open reading frame was identified beginning with an ATG initiation codon at nucleotide 10 and ending with a TAG codon at nucleotide 1006. The encoded protein has 332 amino acid residues. There is 1 amino acid difference for NOV14 with respect to NOV12.

TABLE 91

AGCCTAGAAATGTTAGAGGGTGTTGAGCATCTCCTTCTGCTACTTCTTTTGACAGATGTGAACAGCAAGG
AACTGCAAAGTGGAAACCAGACTTCTGTGTCTCACTTCATTTTGGTGGGCCTGCACCACCCACCACAGCT
GGGAGCGCCACTCTTCTTAGCTTTCCTTGTCATCTATCTCCTCACTGTTTCTGGAAATGGGCTCATCATC
CTCACTGTCTTAGTGGACATCCGGCTCCATCGTCCCATGTGCTTGTTCCTGTGTCACCTCTCCTTCTTGG
ACATGACCATTTCTTGTGCTATTGTCCCCAAGATGCTGGCTGGCTTTCTCTTGGGTAGTAGGATTATCTC
CTTTGGGGGCTGTGTAATCCAACTATTTTCTTTCCATTTCCTGGGCTGTACTGAGTGCTTCCTTTACACA
CTCATGGCTTATGACCGTTTCCTTGCCATTTGTAAGCCCTTACACTATGCTACCATCATGACCCACAGAG
TCTGTAACTCCCTGGCTTTAGGCACCTGGCTGGGAGGGACTATCCATTCACTTTTCCAAACAAGTTTTGT
ATTCCGGCTGCCCTTCTGTGGCCCCAATCGGGTCGACTACATCTTCTGTGACATTCCTGCCATGCTGCGT
CTAGCCTGCGCCGATACGGCCATCAACGAGCTGGTCACCTTTGCAGACATTGGCTTCCTGGCCCTCACCT
GCTTCATGCCCATCCTCACTTCCTATGGCTATATTGTAGCTGCCATCCTGCGAATTCCGTCAGCAGATGG
GCGCCGCAATGCCTTCTCCACTTGTGCTGCCCACCTCACTGTTGTCATTGTTTACTATGTGCCCTGCACC
TTCATTTACCTGCGGCCTTGTTCACAGGAGCCCCTGGATGGGGTGGTAGCTGTCTTTTACACTGTCATCA
CTCCCTTGCTTAACTCCATCATCTACACACTGTGCAACAAAGAAATGAAGGCAGCATTACAGAGGCTAGG
GGGCCACAAGGAAGTGCAGCCTCACTGACTTCAT SEQ ID NO:27

TABLE 92

MLEGVEHLLLLLLLTDVNSKELQSGNQTSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTV
LVDIRLHRPMCLFLCHLSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMA
YDRFLAICKPLHYATIMTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCDIPAMLRLAC
ADTAINELVTFADIGFLALTCFMPILTSYGYIVAAILRIPSADGRRNAFSTCAAHLTVVIVYYVPCTFIY
LRPCSQEPLDGVVAVFYTVITPLLNSIIYTLCNKEMKAALQRLGGHKEVQPH SEQ ID NO:28

In a search of sequence databases, it was found, for example, that NOV14 (SEQ ID NO:27; Query) was found to have 184 of 297 amino acid residues (61%) identical to SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus norvegicus (Subject) and 224 of 297 residues (75%) positive with, the 318 amino acid residue protein (Table 93). The full amino acid sequence of NOV14 was found to have 136 of 305 amino acid residues (44%) identical to SPTREMBL-ACC:Q9UGF6 BA150A6.2 (NOVEL 7 TRANSMEMBRANE RECEPTOR

(RHODOPSIN FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN (HS6M1-21)) - Homo sapiens (Human) (Subject), and 201 of 305 residues (65%) positive with, the 321 amino acid residue protein (Table 94). The full amino acid sequence of NOV14 was found to have 184 of 297 amino acid residues (61%) identical to SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus norvegicus(Rat) (Subject), and 224 of 297 residues (75%) positive with, the 318 amino acid residue protein (Table 95).

TABLE 93

```
Best hits (BLASTP Non-Redundant Composite database):
>ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus
norvegicus(Rat), 318 aa.
Top  Previous Match  Next Match
           Length = 318
 Score = 962 (338.6 bits), Expect = 1.2e-96, P = 1.2e-96
 Identities = 184/297 (61%), Positives = 224/297 (75%)

Query:    28 TSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTVLVDIRLH-RPMCLFLCH 86
             T V+ F+L+GL HPP L   LFL FL+IY+LT  GN LI+LTV  D +LH RPM + L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:    87 LSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMAYDRFLA 146
             LSFLDM +S  IVP+++  F   ++ I+FGGCV QL+ FHFLG T+CFLYTLMAYDR+LA
Sbjct:    71 LSFLDMWLSSVIVPRIILNFTPANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYLA 130

Query:   147 ICKPLHYATIMTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCDIPAML 206
             IC+PL Y  +M  ++C  L  G W+ G+IH   Q +  FRLP+CGP  VDY FCDIPA+L
Sbjct:   131 ICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAVL 190

Query:   207 RLACADTAINELVTFADIGFLALTCFMPILTSYGYIVAAILRIPSADGRRNAFSTCAAHL 266
             RLACADTAINELVTF DIG +A +CF+ IL SY  IV AIL+I +ADGRR AFSTC +HL
Sbjct:   191 RLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSHL 250

Query:   267 TVVIVYYVPCTFIYLRPCSQEPLDGVVAVFYTVITPLLNSIIYTLCNKEMKAALQRL 323 SEQ
ID NO:116
             TVV VYYVPC FIYLR  S+   DG VAVFYTV+TPLLN +IYTL N+E+ +AL+RL
Sbjct:   251 TVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRL 307
SEQ ID NO:117
```

TABLE 94

```
Best hit (human sequence) =
>ptnr:SPTREMBL-ACC:Q9UGF6 BA150A6.2 (NOVEL 7 TRANSMEMBRANE RECEPTOR (RHODOPSIN
FAMILY) (OLFACTORY RECEPTOR LIKE) PROTEIN (HS6M1-21)) - Homo sapiens (Human),
321 aa.
Top  Previous Match  Next Match
           Length = 321
 Score = 702 (247.1 bits), Expect = 4.1e-69, P = 4.1e-69
 Identities = 136/305 (44%), Positives = 201/305 (65%)

Query:    22 LQSGNQTSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTVLVDIRLHRPMC 81
             ++  NQT+++ FI++G  +  +L   LF  F + Y  T+ GN LIILT + D  LH PM
```

```
Sbjct:    1 MERKNQTAITEFIILGFSNLNELQFLLFTIFFLTYFCTLGGNILIILTTVTDPHLHTPMY 60

Query:   82 LFLCHLSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMAY 141
            FL +L+F+D+  + + VP+M+   L   + IS+ GCV+QLF+F F   +EC L   MAY
Sbjct:   61 YFLGNLAFIDICYTTSNVPQMMVHLLSKKKSISYVGCVVQLFAFVFFVGSECLLLAAMAY 120

Query:  142 DRFLAICKPLHYATIMTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCD 201
            DR++AIC PL Y+ I++  +CN LA   W  G ++S+  T   F LPFCG N+++Y FCD
Sbjct:  121 DRYIAICNPLRYSVILSKVLCNQLAASCWAAGFLNSVVHTVLTFCLPFCGNNQINYFFCD 180

Query:  202 IPAMLRLACADTAINELVTFADIGFLALTCFMPILTSYGYIVAAILRIPSADGRRNAFST 261
            IP +L L+C +T++NEL   +   F+  T F+ I+ SY  I++ ILRI S++GRR AFST
Sbjct:  181 IPPLLILSCGNTSVNELALLSTGVFIGWTPFLCIVLSYICIISTILRIQSSEGRRKAFST 240

Query:  262 CAAHLTVVIVYYVPCTFIYLRPCSQEPL--DGVVAVFYTVITPLLNSIIYTLCNKEMKAA 319
            CA+HL +V ++Y    F Y+RP S   L  D +V+V Y+V+TP+LN IIYTL NK++K A
Sbjct:  241 CASHLAIVFLFYGSAIFTYVRPISTYSLKKDRLVSVLYSVVTPMLNPIIYTLRNKDIKEA 300

Query:  320 LQRLG 324  SEQ ID NO:118
            ++ +G
Sbjct:  301 VKTIG 305  SEQ ID NO:119
```

TABLE 95

```
BLASTX (non redundant) =
>ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus
norvegicus(Rat), 318 aa.
Top  Previous Match  Next Match
            Length = 318
  Plus Strand HSPs:
 Score = 962 (338.6 bits), Expect = 6.9e-96, P = 6.9e-96
 Identities = 184/297 (61%), Positives = 224/297 (75%), Frame = +1

Query:   91 TSVSHFILVGLHHPPQLGAPLFLAFLVIYLLTVSGNGLIILTVLVDIRLH-RPMCLFLCH 267
            T V+ F+L+GL HPP L   LFL FL+IY+LT  GN LI+LTV  D +LH RPM + L
Sbjct:   11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:  268 LSFLDMTISCAIVPKMLAGFLLGSRIISFGGCVIQLFSFHFLGCTECFLYTLMAYDRFLA 447
            LSFLDM +S  IVP+++  F   ++ I+FGGCV QL+ FHFLG T+CFLYTLMAYDR+LA
Sbjct:   71 LSFLDMWLSSVIVPRIILNFTPANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYLA 130

Query:  448 ICKPLHYATIMTHRVCNSLALGTWLGGTIHSLFQTSFVFRLPFCGPNRVDYIFCDIPAML 627
            IC+PL Y  +M  ++C  L  G W+ G+IH   Q +  FRLP+CGP  VDY FCDIPA+L
Sbjct:  131 ICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAVL 190

Query:  628 RLACADTAINELVTFADIGFLALTCFMPILTSYGYIVAAILRIPSADGRRNAFSTCAAHL 807
            RLACADTAINELVTF DIG +A +CF+ IL SY  IV AIL+I +ADGRR AFSTC +HL
Sbjct:  191 RLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSHL 250

Query:  808 TVVIVYYVPCTFIYLRPCSQEPLDGVVAVFYTVITPLLNSIIYTLCNKEMKAALQRL 978
SEQ ID NO:120
            TVV VYYVPC FIYLR  S+   DG VAVFYTV+TPLLN +IYTL N+E+ +AL+RL
Sbjct:  251 TVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRL 307
SEQ ID NO:121
```

Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (certainity =0.6400). The polypeptide seems to have a cleavable N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 19 and 20.

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs (Table 96).

TABLE 96

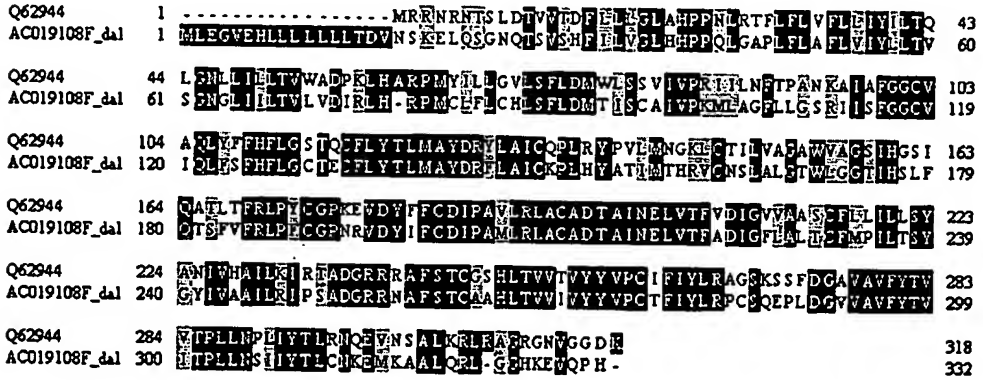
69: T->C(6)
118765007(i), phred 23
118765027(i), phred 33
118765030(i), phred 33
118765038(i), phred 33
118765034(i), phred 30
118878829(i), phred 21
80: T->C(2)
118765582(i), phred 41
118765588(i), phred 49
111: T->G(2)
126554869(i), phred 37
126727189(i), phred 49
205: C->T(2)
118765582(i), phred 26
118765588(i), phred 33
300: C->T(2)
118765582(i), phred 34
118765588(i), phred 45
437: G->C(8)
126554851(i), phred 29
126554969(i), phred 31
126554984(i), phred 35
126727421(i), phred 36
126727378(i), phred 38
126727158(i), phred 49
118765036(i), phred 45
126554869(i), phred 40
525: C->gap(3)
126608134(i), phred 123
118765030(i), phred 123
118765588(i), phred 123
547: T->gap(2)
126608134(i), phred 123
118765577(i), phred 123
738: C->T(4)
118765025(i), phred 33
126554926(i), phred 36
118878831(i), phred 36
118878843(i), phred 27
766: A->G(2)
118877422(i), phred 28

118877416(i), phred 26
865: C->T(2)
126554926(i), phred 49
126727273(i), phred 49
963: T->A(2)
126554926(i), phred 33
126727273(i), phred 33

A multiple sequence alignment is given in Table 97, with the protein of the invention being shown on line 3, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV14 protein is a novel member of the OR family. In the following table, AC019108_F_da1 is the full-length SEQ ID NO:28. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 97

Multiple Alignment:



NOV15

In the present invention, the target sequence identified previously, NOV6, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at

the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated NOV15. NOV15 is 1038 nucleotides (Table 98) encoding a novel GPCR-like protein (Table 99). An open reading frame was identified beginning with an ATG initiation codon at nucleotide 28 and ending with a TAG codon at nucleotide 961. The encoded protein has 311 amino acid residues. This differs from NOV6 at bp: 349, 356.

TABLE 98

AGGGAGAGAGACCAAGGGTGAGAAGAAATGTCCAACGCCAGCCTACTGACAGCGTTCATCCTCATGGGCC
TTCCCCATGCCCCAGCGCTGGACGCCCCCCTCTTTGGAGTCTTCCTGGTGGTTTACGTGCTCACTGTGCT
GGGGAACCTCCTCATCCTGCTGGTGATCAGGGTGGATTCTCACCTCCACACCACCATGTACTACTTCCTC
ACCAACCTGTCGTTCATTGACATGTGGTTCTCCACTGTCACGGTGCCCAAATTGCTGATGACTTTGGTGT
TCCCAAGTGGCAGGGCTATCTCCTTCCACAGCTGCATGGCTCAGCTCTATTTCTTTCACTTCCTAGGGGG
CACCGAGTGTTTCCTCTACAGGGTCATGTCCTGTGATCGCTACCTGGCCATCAGTTACCCGCTCAGGTAC
ACCAGCATGATGACTGGGCGCTCGTGTACTCTTCTGGCCACCAGCACTTGGCTCAGTGGCTCTCTGCACT
CTGCTGTCCAGGCCATATTGACTTTCCATTTGCCCTACTGTGGACCCAACTGGATCCAGCACTATTTGTG
TGATGCACCGCCCATCCTGAAACTGGCCTGTGCAGACACCTCAGCCATAGAGACTGTCATTTTTGTGACT
GTTGGAATAGTGGCCTCGGGCTGCTTTGTCCTGATAGTGCTGTCCTATGTGTCCATCGTCTGTTCCATCC
TGCGGATCCGCACCTCAGAGGGGAAGCACAGAGCCTTTCAGACCTGTGCCTCCCACTGTATCGTGGTCCT
TTGCTTCTTTGGCCCTGGTCTTTTCATTTACCTGAGGCCAGGCTCCAGGAAAGCTGTGGATGGAGTTGTG
GCCGTTTTCTACACTGTGCTGACGCCCCTTCTCAACCCTGTTGTGTACACCCTGAGGAACAAGGAGGTGA

AGAAAGCTCTGTTGAAGCTGAAAGACAAAGTAGCACATTCTCAGAGCAAATAGACACTAGGGAAGATTAC
ATATCTTAGCTCTTGTGAATAGTGCTGTGAAAAACATACAGGGGCAGGTATCTTTTGG SEQ ID NO:29

TABLE 99

MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYFLTNLSFIDMW
FSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCDRYLAISYPLRYTSMMTGRSC
TLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPILKLACADTSAIETVIFVTVGIVASGCF
VLIVLSYVSIVCSILRIRTSEGKHRAFQTCASHCIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTP
LLNPVVYTLRNKEVKKALLKLKDKVAHSQSK SEQ ID NO:30

In a search of sequence databases, it was found, for example, that NOV15 (Query) was found to have 165 of 298 amino acid residues (55%) identical to SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus norvegicus(Rat) (Subject) and 223 of 298 residues (74%) positive with, the 318 amino acid residue protein (Table 100). The full amino acid sequence of NOV15 was found to have 142 of 312 amino acid residues (45%) identical to SPTREMBL- ACC:O95006 WUGSC:H_DJ0669B10.1 PROTEIN - Homo sapiens (Human) (Subject), and 208 of 312 residues (66%) positive with, the 317 amino acid residue protein (Table 101). The full amino acid sequence of NOV15 (Query) was found to have 165 of 298 amino acid residues (55%) identical to SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus norvegicus(Rat) (Subject), and 223 of 298 residues (74%) positive with, the 318 amino acid residue protein (Table 102).

TABLE 100

```
Best hits (BLASTP Non-Redundant Composite database):
>ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus
norvegicus(Rat), 318 aa.
Top  Previous Match  Next Match
           Length = 318
 Score = 903 (317.9 bits), Expect = 2.1e-90, P = 2.1e-90
 Identities = 165/298 (55%), Positives = 223/298 (74%)

Query:     5 SLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTT-MYYFLTN 63
             +++T F+L+GL H P L   LF VFL++Y+LT LGNLLILL +  D  LH   MY  L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:    64 LSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCDRYL 123
             LSF+DMW S+V VP++++    P+ +AI+F  C+AQLYFFHFLG T+CFLY +M+ DRYL
Sbjct:    71 LSFLDMWLSSVIVPRIILNFT-PANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYL 129

Query:   124 AISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPI 183
             AI  PLRY  +M G+ CT+L    W++GS+H ++QA LTF LPYCGP  + ++ CD P +
Sbjct:   130 AICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAV 189

Query:   184 LKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTCASH 243
             L+LACADT+  E V FV +G+VA+ CF+LI+LSY +IV +IL+IRT++G+ RAF TC SH
Sbjct:   190 LRLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSH 249
```

```
Query:   244 CIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNKEVKKALLKLK 302
SEQ ID NO:123
              VV  ++ P +FIYLR GS+ + DG VAVFYTV+TPLLNP++YTLRN+EV  AL +L+
Sbjct:   250 LTVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRLR 308
SEQ ID NO:124
```

TABLE 101

```
Best hit (human sequence) =
>ptnr:SPTREMBL-ACC:O95006 WUGSC:H_DJ0669B10.1 PROTEIN - Homo sapiens
(Human),317 aa.
Top  Previous Match  Next Match
           Length = 317
 Score = 694 (244.3 bits), Expect = 2.9e-68, P = 2.9e-68
 Identities = 142/312 (45%), Positives = 208/312 (66%)

Query:     1 MSNASLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTTMYYF 60
             + N + +  FIL+GL          LF +FLV Y++TVLGN LI+L+IR+DS LHT MY+F
Sbjct:     3 IDNQTWVREFILLGLSSDWCTQISLFSLFLVTYLMTVLGNCLIVLLIRLDSRLHTPMYFF 62

Query:    61 LTNLSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCD 120
             LTNLS +D+ ++T  VP+LL   +    +AI F SC AQL+F   LGG E  L  VM+ D
Sbjct:    63 LTNLSLVDVSYATSVVPQLLAHFL-AEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAYD 121

Query:   121 RYLAISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDA 180
             R++A+S  LRY+++M G  C  LA ++W+SGS++S VQ  +TF LP C   +I H  C+
Sbjct:   122 RHVAVSDRLRYSAIMHGGLCARLAITSWVSGSINSLVQTAITFQLPMCTNKFIDHISCEL 181

Query:   181 PPILKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTC 240
               +++LAC DTS+ E . I V+  ++    F L++LSY+ I+ +IL+I++ EG+ +AF TC
Sbjct:   182 LAVVRLACVDTSSNEAAIMVSSIVLLMTPFCLVLLSYIRIISTILKIQSREGRKKAFHTC 241

Query:   241 ASHCIVVLCFFGPGLFIYLRPGSRKAV--DGVVAVFYTVLTPLLNPVVYTLRNKEVKKAL 298
             ASH  VV   +G  +F Y++P S  +V  + +++VFY ++ PLLNPV+Y+LRNKEVK A
Sbjct:   242 ASHLTVVALCYGTTIFTYIQPHSGPSVLQEKLISVFYAIVMPLLNPVIYSLRNKEVKGAW 301

Query:   299 LKLKDKVAHSQSK 311  SEQ ID NO:125
              KL +K +   SK
Sbjct:   302 HKLLEKFSGLTSK 314  SEQ ID NO:126
```

TABLE 102

```
>ptnr:SPTREMBL-ACC:Q62944 TASTE BUD RECEPTOR PROTEIN TB 641 - Rattus
norvegicus(Rat), 318 aa.
Top  Previous Match  Next Match
           Length = 318
  Plus Strand HSPs:
 Score = 903 (317.9 bits), Expect = 1.2e-89, P = 1.2e-89
 Identities = 165/298 (55%), Positives = 223/298 (74%), Frame = +1

Query:    40 SLLTAFILMGLPHAPALDAPLFGVFLVVYVLTVLGNLLILLVIRVDSHLHTT-MYYFLTN 216
             +++T F+L+GL H P L   LF VFL++Y+LT LGNLLILL +  D  LH   MY  L
Sbjct:    11 TVVTDFLLLGLAHPPNLRTFLFLVFLLIYILTQLGNLLILLTVWADPKLHARPMYILLGV 70

Query:   217 LSFIDMWFSTVTVPKLLMTLVFPSGRAISFHSCMAQLYFFHFLGGTECFLYRVMSCDRYL 396
```

```
              LSF+DMW S+V VP++++    P+ +AI+F  C+AQLYFFHFLG T+CFLY +M+ DRYL
Sbjct:    71 LSFLDMWLSSVIVPRIILNFT-PANKAIAFGGCVAQLYFFHFLGSTQCFLYTLMAYDRYL 129

Query:   397 AISYPLRYTSMMTGRSCTLLATSTWLSGSLHSAVQAILTFHLPYCGPNWIQHYLCDAPPI 576
              AI  PLRY  +M G+ CT+L    W++GS+H ++QA LTF LPYCGP  + ++ CD P +
Sbjct:   130 AICQPLRYPVLMNGKLCTILVAGAWVAGSIHGSIQATLTFRLPYCGPKEVDYFFCDIPAV 189

Query:   577 LKLACADTSAIETVIFVTVGIVASGCFVLIVLSYVSIVCSILRIRTSEGKHRAFQTCASH 756
              L+LACADT+  E V FV +G+VA+ CF+LI+LSY +IV +IL+IRT++G+ RAF TC SH
Sbjct:   190 LRLACADTAINELVTFVDIGVVAASCFLLILLSYANIVHAILKIRTADGRRRAFSTCGSH 249

Query:   757 CIVVLCFFGPGLFIYLRPGSRKAVDGVVAVFYTVLTPLLNPVVYTLRNKEVKKALLKLK 933
SEQ ID NO:127
                VV  ++ P +FIYLR GS+ + DG VAVFYTV+TPLLNP++YTLRN+EV  AL +L+
Sbjct:   250 LTVVTVYYVPCIFIYLRAGSKSSFDGAVAVFYTVVTPLLNPLIYTLRNQEVNSALKRLR 308
SEQ ID NO:128
```

Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (certainity =0.6400). The polypeptide seems to have a cleavable N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 51 and 52. NOV 15 differs from NOV 9 at (bp): 349, 356 and (aa):308, 309, 310. Changed bp 110, 381-383, 388-390, 404-405 b/c of phrap to match clone

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs (Table 103).

TABLE 103

```
Cons.Pos.: 132  Depth:  10  Change: C > T
    Putative Allele Freq.:  0.200
  -> 118878822(+,i)  unrev. Fpos: 119
  -> 129054605(+,i)  unrev. Fpos: 152

    Cons.Pos.: 144  Depth:  10  Change: C > T
    Putative Allele Freq.:  0.200
  -> 118878822(+,i)  unrev. Fpos: 131
  -> 129054605(+,i)  unrev. Fpos: 164

    Cons.Pos.: 170  Depth:  10  Change: G > A
    Putative Allele Freq.:  0.200
  -> 118878822(+,i)  unrev. Fpos: 157
  -> 129054605(+,i)  unrev. Fpos: 190

    Cons.Pos.: 174  Depth:  10  Change: G > A
    Putative Allele Freq.:  0.200
  -> 118878822(+,i)  unrev. Fpos: 161
  -> 129054605(+,i)  unrev. Fpos: 194

    Cons.Pos.: 177  Depth:  10  Change: C > A
    Putative Allele Freq.:  0.200
```

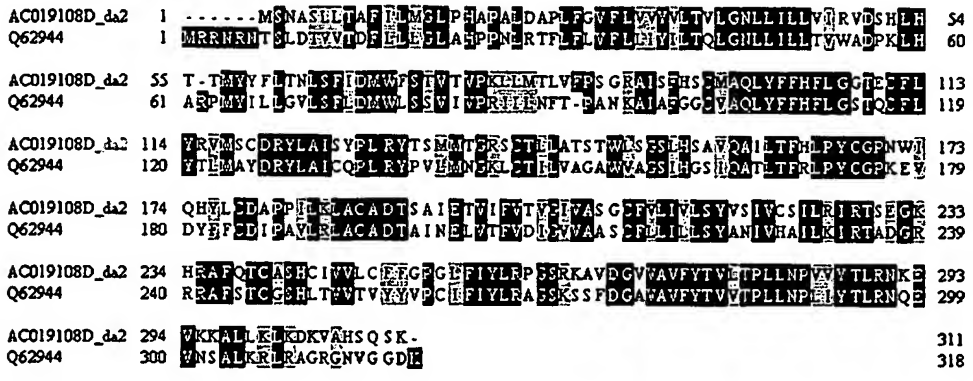
```
-> 118878822(+,i)  unrev. Fpos: 164
-> 129054605(+,i)  unrev. Fpos: 197

  Cons.Pos.: 184  Depth:  10  Change: A > G
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 171
-> 129054605(+,i)  unrev. Fpos: 204

  Cons.Pos.: 185  Depth:  10  Change: C > T
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 172
-> 129054605(+,i)  unrev. Fpos: 205

  Cons.Pos.: 186  Depth:  10  Change: A > G
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 173
-> 129054605(+,i)  unrev. Fpos: 206

  Cons.Pos.: 207  Depth:  10  Change: T > C
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 194
-> 129054605(+,i)  unrev. Fpos: 227

  Cons.Pos.: 221  Depth:  10  Change: T > C
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 208
-> 129054605(+,i)  unrev. Fpos: 241

  Cons.Pos.: 316  Depth:  10  Change: T > C
  Putative Allele Freq.:  0.200
-> 118878822(+,i)  unrev. Fpos: 303
-> 129054605(+,i)  unrev. Fpos: 336

  Cons.Pos.:1026  Depth:   9  Change: - > C
  Putative Allele Freq.:  0.222
-> 129054554(-,i)  unrev. Fpos:  91
-> 129054568(-,i)  unrev. Fpos:  92
```

A multiple sequence alignment is given in Table 104, with the protein of the invention being shown on line 1, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV15 protein is a novel member of the OR family. In the following table, AC019108D_da2 is

the full-length SEQ ID NO:8. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 104

Multiple Alignment:



NOV16

NOV16 is 995 nucleotides (SEQ ID NO: 31) (designated CuraGen Acc. No. nh0413n10_da4) and encodes a novel Olfactory Receptor -like protein (SEQ ID NO:32) is shown in Table 105. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 31-33 and ending with a TGA codon at nucleotides 973-75. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 105, and the start and stop codons are in bold letters. The encoded protein having 314 amino acid residues is presented using the one-letter code in Table 106.

TABLE 105

TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCT
TTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGC
TCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATCTCTCCTTTGCAGAT
CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTG
TATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCCGTGATGTCCTATGACCGGTATG
TGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCC
AGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTA
CTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCG
TGGTAATCCTCCTGGCCCCTATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCT
GGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATT
CACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAA
TGTTGAACCCCATAATTTATAGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTC
TCTCATAGGCAGTGACCTCTGAGTCTGACTTTTAA SEQ ID NO:31

TABLE 106

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLRNLSFADLCFSTSIV
PQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVS
LVDTSFTFHLPYWGQNIINHYFCEPPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGR
LKAFSTCGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVGRKCFSH
RQ SEQ ID NO:32

In a search of sequence databases, it was found, for example, that NOV16 (Query) has 182 of 216 bases (84%) identical to TREMBLNEW-ACC:AAD43436 OLFACTORY RECEPTOR - Mus musculus domesticus (western European house mouse) (Subject) (Table 107), and 218 of 307 residues (71%) positive with, the 216 amino acid protein. NOV16 (Query) has 166 of 305 bases (54%) identical to SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens (Human) (Subject), and 217 of 305 residues (71%) positive with, the 217 amino acid protein. (Table 108).

TABLE 107

ptnr:TREMBLNEW-ACC:AAD43436 OLFACTORY RECEPTOR - Mus musculus domesticus (western European house mouse), 216 aa (fragment).

Top Previous Match Next Match
Length = 216

Score = 941 (331.2 bits), Expect = 2.0e-94, P = 2.0e-94
Identities = 182/216 (84%), Positives = 194/216 (89%)

```
Query:    68 FADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVC 127
             FADLCFST+ VPQVLVHFLVKRKTISF GC TQI+V LLVGCTECALLAVMSYDRYVAVC
Sbjct:     1 FADLCFSTTTVPQVLVHFLVKRKTISFAGCSTQIVVLLLVGCTECALLAVMSYDRYVAVC 60

Query:   128 KPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKL 187
             KPL+YSTIMT  +C+ L+  SWASGALVSLVDT+FT  LPY G N+INH+FCEPPALLKL
Sbjct:    61 KPLHYSTIMTHWLCVQLAAGSWASGALVSLVDTTFTLRLPYRGNNVINHFFCEPPALLKL 120

Query:   188 ASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIV 247
             AS DTYSTEMAIF+MGVVILLAP+SLIL SYWNIISTVIQMQSGEGRLK FSTCGSHLIV
Sbjct:   121 ASADTYSTEMAIFAMGVVILLAPVSLILTSYWNIISTVIQMQSGEGRLKVFSTCGSHLIV 180
SEQ ID NO:131

Query:   248 VVLFYGSGIFTYMRPNSKTTK  SEQ ID NO:130
```

TABLE 108

>ptnr:SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens (Human), 317 aa.

Top Previous Match Next Match
Length = 317

```
 Score = 842 (296.4 bits), Expect = 6.0e-84, P = 6.0e-84
 Identities = 166/305 (54%), Positives = 217/305 (71%)

Query:     1 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 60
             MG +NQT+VS+FI LGLS D  T++ LF+LFL++Y++TVLGN LI++LI LDSRLHTPMY
Sbjct:     1 MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60

Query:    61 FFLRNLSFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 120
             FFL NLS  D+ ++TS+VPQ+L HFL + K I F  C  Q+   L +G  E  LLAVM+Y
Sbjct:    61 FFLTNLSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120

Query:   121 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 180
             DRYVAVC  L YS IM   +C  L+  SW SG + S V T+ TF LP       I+H  CE
Sbjct:   121 DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180
SEQ ID NO:133

Query:   181 PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQS SEQ ID NO:132
```

Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (certainity =0.6000). The polypeptide seems to have NO N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 41 and 42.

NOV17

The novel nucleic acid of 994 nucleotides (SEQ ID NO:33) (designated CuraGen Acc. No. CG55604-06) encoding a novel Olfactory Receptor-like protein (SEQ ID NO:34) is shown in Table 109. An open reading frame was identified beginning with an ATG initiation codon at nucleotides 31-33 and ending with a TGA codon at nucleotides 973-975. Putative untranslated regions, if any, are found upstream from the initiation codon and downstream from the termination codon. The encoded protein having 314 amino acid residues is presented using the one-letter code in Table 110.

TABLE 109

```
TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCC   60
AAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTT  120
TTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTC  180
CTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGAT  240
CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAA  300
ACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACA  360
GAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTG  420
TACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCC  480
AGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGA  540
CAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATA  600
GACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCT  660
GTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCT  720
```

```
GGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTC   780
TTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTG   840
GATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTAT   900
AGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTC   960
TCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA SEQ ID NO:33
994
```

TABLE 110

```
MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY    60
FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY   120
DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE   180
PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST   240
CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA   300
LRKLVGRKCFSHRQ  SEQ ID NO:34                                    314
```

The sequence of NOV17 is highly similar to that of NOV13. A sequence difference results in a change from Trp249 to Ser249 (W249S). In a search of sequence databases, it was found, for example, that NOV17 (Query) has 606 of 921 bases (65%) identical to a gb:GENBANK-ID:HSU56421|acc:U56421.1 mRNA from Homo sapiens (Human olfactory receptor (OLF3) gene, complete cds) (Subject) (Table 111 and 113). The full amino acid sequence of NOV17 (Query) was found to have 254 of 304 amino acid residues (83%) identical to, and 275 of 304 amino acid residues (90%) similar to, the 307 amino acid residue ptnr:TREMBLNEW-ACC:AAG45200 protein from Mus musculus (Mouse) (B5 OLFACTORY RECEPTOR) (Subject) (Table 112 and 113).

TABLE 111

```
>gb:GENBANK-ID:HSU56421|acc:U56421.1 Human olfactory receptor (OLF3) gene,
           complete cds - Homo sapiens, 954 bp.
           Length = 954

  Plus Strand HSPs:

 Score = 1588 (238.3 bits), Expect = 9.9e-66, P = 9.9e-66
 Identities = 606/921 (65%), Positives = 608/921 (66%), Strand = Plus / Plus

Query:    31 ATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAG-GA 89
             ATGGGA  AGA AACCA AC T  GTG    A TTTAT  TCCT GG CT TC CAG GA
Sbjct:     1 ATGGGAACAGATAACCAGACTTGGGTGAGTGAATTTATTCTCCTCGGCCTGTC-CAGTGA 59

Query:    90 CTTGCAGACCCAGATCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCT 149
             CT G A AC C G TC   CT TTT TCCT TTC T  TCAT TAT TG TGACCGTGCT
Sbjct:    60 CTGGGACACTCGGGTCTCCCTGTTTGTCCTGTTCTTGGTCATGTATGTGGTGACCGTGCT 119

Query:   150 TGGAAACCAGCTCATCATCATTCTCATCTTCCTGGATTCTCGCCTTCACACTCCCATGTA 209
              GG AAC   CTCAT  TC TTCT ATC   CTGGA    CG CT CACACTCCCATGTA
Sbjct:   120 GGGGAACTGTCTCATTGTCCTTCTGATCAGACTGGACAGCCGACTCCACACTCCCATGTA 179
```

```
Query:  210 TTTTTTTCTTAGAAATATCTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGTCCCTCA 269
            TTT TTTCT A  AA  TCTCC TTG  GAT TCT  T   C AC AG  T GTCCCTCA
Sbjct:  180 TTTCTTTCTCACCAACCTCTCCCTTGTCGATGTCTCCTATGCCACAAGTGTAGTCCCTCA 239

Query:  270 AGTGTTGGTTCACTT-CTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTGTATGACAC 328
              TG TGG  CA TT CTTG   AA A  AAA CCAT  C TT  A  G TGT    C C
Sbjct:  240 GCTGCTGGCACATTTTCTTGCAGAACAT-AAAGCCATCCCATTCCAGAGCTGTGCAGCCC 298

Query:  329 AGATAATTGTCTTTCTTCTGGTTGGGTGT-ACAGAGTGTGCGCTGCTGGCAGTGATGTCC 387
            AG TA TT TCT  CT     TTGGGTG  A  GAGT TG  CT CT+GC GTGATG CC
Sbjct:  299 AGTTATTTTTCTCCCTGGCC-TTGGGTGGGATTGAGTTTGTTCTCCTSGCGGTGATGGCC 357

Query:  388 TATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTACTACTCTACCATCATGACACAACGG 447
            TATGACCG TATGTGGCTGT TG  A  CCCT+   TACTC  CCATCATG CA    GG
Sbjct:  358 TATGACCGCTATGTGGCTGTGTGTGATGCCCTSCGATACTCGGCCATCATG-CATGGAGG 416

Query:  448 G-TGTGTCTCT-GGCTGTCCTTCAGGTCCTGGGCCAGTGGGG-CACTAGTGTCTTTAGTA 504
            G TGTGT  CT GG TG CC TCA  TCCTGGG CAGTGG   CA  AG  TCT   GT
Sbjct:  417 GCTGTGTG-CTAGGTTGGCCATCACATCCTGGGTCAGTGGCTTCATCAGC-TCTCCTGTG 474

Query:  505 GATACCAGCTTT-ACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCAC-TACT 562
             A AC  GCT T AC TT CA CT CCC   TG  GA A AA  T AT  ATCAC TA T
Sbjct:  475 CAGACT-GCTATCACCTTTCAGCTGCCCATGTGCAGAAACAAGTTTATTGATCACATA-T 532

Query:  563 TTTGTGAACCTCCT-GCCCTCCTGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATG 621
              TGTGAAC TCCT GC  T  T A GCTGGCTT   T GACAC T C  CA  GA  T
Sbjct:  533 CCTGTGAAC-TCCTAGCTGTGGTCAGGCTGGCTTGTGTGGACACCTCCTCCAATGAGGTC 591

Query:  622 GCCATCTTT-TCAATGGGCGTGGTAATCCTCCTGGCCCCTGTCTCCCTGATTCTTGGTTC 680
             CCATC T  T   T G C T GT  T CT  TG C CC  TCT CCTG TTCTT   TC
Sbjct:  592 ACCATCATGGTGTCTAG-CATTGTTCTTCTGATGACACCCCTCTGCCTGGTTCTTTTGTC 650

Query:  681 TTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGC 740
             TA     A AT ATCTCCAC  T  T  AGAT CAGTC  G GAAGG AGA   AA GC
Sbjct:  651 CTACATCCAGATCATCTCCACCATCCTAAAGATCCAGTCCAGAGAAGGAAGAAAGAAAGC 710

Query:  741 TTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGA-ATAT 799
            TTT   CAC TGTG CTC CA CT A  GT GTTG CCT T CTATGG T  GG  AT T
Sbjct:  711 TTTCCACACGTGTGCCTCTCACCTCACAGTGGTTGCCCTGTGCTATGG-TGTGGCCATTT 769

Query:  800 TCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTGGATAAAATGATATCTGTGT 859
            TCAC TACAT C  CC  ACTCCA   C  C       C GGA AA  TG T TCTGT T
Sbjct:  770 TCACTTACATCCAGCCCCACTCCAGTCCCTCTGTCCTTCAGGAGAAGTTGTTCTCTGTCT 829

Query:  860 TCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGCTTGAGGAACAAAGATG 919
            T TAT C     T AC CCAATG TGAACCCCAT ATTTA AGC T AGGAA AAAGA G
Sbjct:  830 TTTATGCCATTTTAACACCAATGCTGAACCCCATGATTTACAGCCTAAGGAATAAAGAGG 889

Query:  920 TCAAAGGGGC-TCTCAGGAAACTAGTTGGGAGA 951  SEQ ID NO:134
            T AA GGGGC T  CAG AAACTA T  GGA A
Sbjct:  890 TGAAGGGGGCCTGGCAG-AAACTATTATGGAAA 921  SEQ ID NO:135
```

TABLE 112

```
>ptnr:TREMBLNEW-ACC:AAG45200 B5 OLFACTORY RECEPTOR - Mus musculus (Mouse), 307
aa.
           Length = 307

 Score = 1309 (460.8 bits), Expect = 2.3e-133, P = 2.3e-133
 Identities = 254/304 (83%), Positives = 275/304 (90%)

Query:     1 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 60
             MGE+N+T V++FIFLGLSQD QTQ+LLF LFL IYLLTVLGN LII+LI  D RLHTPMY
Sbjct:     1 MGEDNRTSVTEFIFLGLSQDPQTQVLLFFLFLFIYLLTVLGNLLIIVLIHSDPRLHTPMY 60

Query:    61 FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 120
             FFLRN+SFADLCFST+ VPQVLVHFLVKRKTISF GC TQI+V LLVGCTECALLAVMSY
Sbjct:    61 FFLRNLSFADLCFSTTTVPQVLVHFLVKRKTISFAGCSTQIVVLLLVGCTECALLAVMSY 120

Query:   121 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 180
             DRYVAVCKPL+YSTIMT  VC+ L+  SWASGALVSLVDT+FT  LPY G N+INH+FCE
Sbjct:   121 DRYVAVCKPLHYSTIMTHWVCVQLAAGSWASGALVSLVDTTFTLRLPYRGNNVINHFFCE 180

Query:   181 PPALLKLASIDTYSTEMAIFSMGVVILLAPVSLILGSYWNIISTVIQMQSGEGRLKAFST 240
             PPALLKLAS DTYSTEMAIF+MGVVILLAPVSLIL SYWNI+STVIQMQSGEGRLK FST
Sbjct:   181 PPALLKLASADTYSTEMAIFAMGVVILLAPVSLILTSYWNIVSTVIQMQSGEGRLKVFST 240

Query:   241 CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA 300
             CGSHLIVVVLFYGSGIF YMRPNSK   E DKMISVFY+AVTPMLNPIIYSLRNKDVKGA
Sbjct:   241 CGSHLIVVVLFYGSGIFAYMRPNSKIMNEKDKMISVFYSAVTPMLNPIIYSLRNKDVKGA 300

Query:   301 LRKL 304  SEQ ID NO:136
             L+++
Sbjct:   301 LKRI 304  SEQ ID NO:137
```

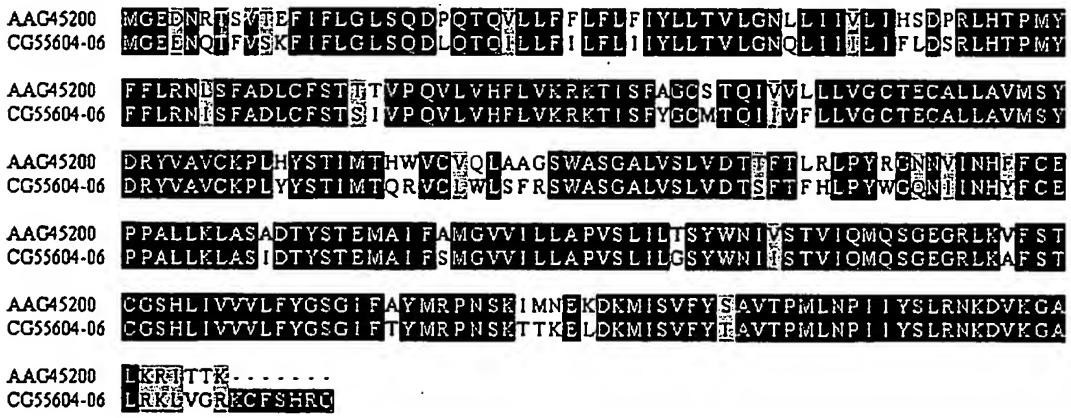
TABLE 113

```
>s3aq:151668831 205 frag (205 non-5'sig-CG), 995 bp.
           Length = 995

  Plus Strand HSPs:

 Score = 4961 (744.3 bits), Expect = 7.1e-219, P = 7.1e-219
 Identities = 993/994 (99%), Positives = 993/994 (99%), Strand = Plus / Plus

Query:     1 TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:     1 TGCCAAACAGGTAAACAGGCAAAAATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCC 60

Query:    61 AAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:    61 AAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGATCCTGCTATTTATCCTT 120

Query:   121 TTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:   121 TTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTC 180
```

```
Query:  181 CTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGAT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  181 CTGGATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGAT 240

Query:  241 CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  241 CTCTGTTTCTCTACTAGCATTGTCCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAA 300

Query:  301 ACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  301 ACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTCTTCTGGTTGGGTGTACA 360

Query:  361 GAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTG 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  361 GAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTG 420

Query:  421 TACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCC 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  421 TACTACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCC 480

Query:  481 AGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  481 AGTGGGGCACTAGTGTCTTTAGTAGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGA 540

Query:  541 CAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  541 CAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCCTGAAGCTGGCTTCCATA 600

Query:  601 GACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCT 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  601 GACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCT 660

Query:  661 GTCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCT 720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  661 ATCTCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCT 720

Query:  721 GGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTC 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  721 GGGGAAGGGAGACTCAAGGCTTTTTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTC 780

Query:  781 TTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  781 TTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCAAGACTACAAAAGAACTG 840

Query:  841 GATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTAT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  841 GATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTAT 900

Query:  901 AGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct:  901 AGCTTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTC 960

Query:  961 TCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA 994 SEQ ID NO:138
            |||||||||||||||||||||||||||||||||
Sbjct:  961 TCTCATAGGCAGTGACCTCTGAGTCTGACTTTTA 994  SEQ ID NO:139
```

A multiple sequence alignment is given in Table 114, with the protein of the invention being shown on the first line in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on its relatedness to members of the GPCR family, the NOV17 protein is a novel member of the OR family. In the following table, CG55604-06 is the full-length SEQ ID NO:34. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 114



The presence of identifiable domains in the protein disclosed herein was determined by searches using algorithms such as PROSITE, Blocks, Pfam, ProDomain, Prints and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/interpro/). The results indicate that this protein contains the following protein domains (as defined by Interpro) at the indicated positions: domain name 7tm_1 (7 transmembrane receptor (rhodopsin family)) at amino acid positions 41 to 290. This indicates that the sequence of the invention has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Tissue expression

The Olfactory Receptor disclosed in this invention is expressed in at least the following tissues: Apical microvilli of the retinal pigment epithelium, arterial (aortic), basal forebrain, brain, Burkitt lymphoma cell lines, corpus callosum, cardiac (atria and ventricle), caudate nucleus, CNS and peripheral tissue, cerebellum, cerebral cortex, colon, cortical neurogenic cells, endothelial (coronary artery and umbilical vein) cells, palate epithelia, eye, neonatal eye, frontal

cortex, fetal hematopoietic cells, heart, hippocampus, hypothalamus, leukocytes, liver, fetal liver, lung, lung lymphoma cell lines, fetal lymphoid tissue, adult lymphoid tissue, Those that express MHC II and III nervous, medulla, subthalamic nucleus, ovary, pancreas, pituitary, placenta, pons, prostate, putamen, serum, skeletal muscle, small intestine, smooth muscle (coronary artery in aortic) spinal cord, spleen, stomach, taste receptor cells of the tongue, testis, thalamus, and thymus tissue. This information was derived by determining the tissue sources of the sequences that were included in the invention including but not limited to SeqCalling sources, Public EST sources, Literature sources, and/or RACE sources.

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs. "Depth" rerepresents the number of clones covering the region of the SNP. The Putative Allele Frequency (Putative Allele Freq.) is the fraction of all the clones containing the SNP. A dash ("-"), when shown, means that a base is not present. The sign ">" means "is changed to".

Cons.Pos.: 366 Depth: 20 Change: G > A
Putative Allele Freq.: 0.100
Cons.Pos.: 447 Depth: 29 Change: G > C
Putative Allele Freq.: 0.069

Cellular Localization and Sorting

The Psort and Hydropathy profile for the Olfactory Receptor-like protein are shown in Fig. 5. The results predict that this sequence has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0.6000. The first 41 amino acids are more likely to be cleaved as a signal peptide based on the SignalP result (Table 115).

TABLE 115

plasma membrane --- Certainty=0.6000(Affirmative) < succ>
Golgi body --- Certainty=0.4000(Affirmative) < succ>
endoplasmic reticulum (membrane) --- Certainty=0.3000(Affirmative) < succ>
microbody (peroxisome) --- Certainty=0.3000(Affirmative) < succ>
INTEGRAL Likelihood =-13.32 Transmembrane 25 - 41 (20 - 53)

INTEGRAL Likelihood = -9.18 Transmembrane 199 - 215 (193 - 222)

INTEGRAL Likelihood = -8.86 Transmembrane 101 - 117 (96 - 120)

INTEGRAL Likelihood = -1.75 Transmembrane 71 - 87 (71 - 87)

INTEGRAL Likelihood = -1.65 Transmembrane 241 - 257 (241 - 257)

Is the sequence a signal peptide?

# Measure	Position	Value	Cutoff	Conclusion
max. C	42	1.000	0.37	YES
max. Y	42	0.629	0.34	YES
max. S	35	0.979	0.88	YES
mean S	1-41	0.593	0.48	YES

Most likely cleavage site between pos. 41 and 42: VLG-NQ

NOV18

In the present invention, the target sequence identified previously, NOV13, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an

assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. NOV18 is 968 nucleotides (SEQ ID NO:35) (Table 114) encoding a novel GPCR-like protein (SEQ ID NO:36) (Table 115). An open reading frame was identified beginning with an ATG initiation codon at nucleotide 8 and ending with a TAG codon at nucleotide 950. The encoded protein has 314 amino acid residues.

TABLE 114

TATATCAATGGGAGAAGAAAACCAAACCTTTGTGTCCAAGTTTATCTTCCTGGGTCTTTCACAGGACTTGCAGACCCAGA
TCCTGCTATTTATCCTTTTCCTCATCATTTATCTGCTGACCGTGCTTGGAAACCAGCTCATCATCATTCTCATCTTCCTG
GATTCTCGCCTTCACACTCCCATGTATTTTTTTCTTAGAAATATCTCCTTTGCAGATCTCTGTTTCTCTACTAGCATTGT
CCCTCAAGTGTTGGTTCACTTCTTGGTAAAGAGGAAAACCATTTCTTTTTATGGGTGTATGACACAGATAATTGTCTTTC
TTCTGGTTGGGTGTACAGAGTGTGCGCTGCTGGCAGTGATGTCCTATGACCGGTATGTGGCTGTCTGCAAGCCCCTGTAC
TACTCTACCATCATGACACAACGGGTGTGTCTCTGGCTGTCCTTCAGGTCCTGGGCCAGTGGGGCACTAGTGTCTTTAGT
AGATACCAGCTTTACTTTCCATCTTCCCTACTGGGGACAGAATATAATCAATCACTACTTTTGTGAACCTCCTGCCCTCC
TGAAGCTGGCTTCCATAGACACTTACAGCACAGAAATGGCCATCTTTTCAATGGGCGTGGTAATCCTCCTGGCCCCTATC
TCCCTGATTCTTGGTTCTTATTGGAATATTATCTCCACTGTTATCCAGATGCAGTCTGGGGAAGGGAGACTCAAGGCTTT
TTCCACCTGTGGCTCCCATCTTATTGTTGTTGTCCTCTTCTATGGGTCAGGAATATTCACCTACATGCGACCAAACTCCA
AGACTACAAAAGAACTGGATAAAATGATATCTGTGTTCTATACAGCGGTGACTCCAATGTTGAACCCCATAATTTATAGC
TTGAGGAACAAAGATGTCAAAGGGGCTCTCAGGAAACTAGTTGGGAGAAAGTGCTTCTCTCATAGGCAGTGACCTCTGAG
TCTGACTA (SEQ ID NO:140)

TABLE 115

MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMYFFLR
NISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVA
VCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALL
KLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIV
VVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGALRKLVG
RKCFSHRQ (SEQ ID NO:141)

In a search of sequence databases, it was found, for example, that NOV18 (Query) was found to have 182 of 216 amino acid residues (84%) identical to, and 194 of 216 residues (89%) positive with TREMBLNEW-ACC:AAD43436 OLFACTORY RECEPTOR - Mus musculus domesticus (western European house mouse) (Subject) (Table 116). NOV18 (Query) was found to have 165 of 305 amino acid residues (54%) identical to, and 217 of 305 residues (71%)

positive with, SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens (Human) (Subject) (Table 117).

TABLE 116

```
Best hits (BLASTP Non-Redundant Composite database):
     ptnr:TREMBLNEW-ACC:AAD43436 OLFACTORY RECEPTOR - Mus musculus domesticus
(western European house mouse), 216 aa (fragment).

Top  Previous Match  Next Match
           Length = 216

 Score = 941 (331.2 bits), Expect = 1.9e-94, P = 1.9e-94
 Identities = 182/216 (84%), Positives = 194/216 (89%)

Query:    68 FADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSYDRYVAVC 127
             FADLCFST+ VPQVLVHFLVKRKTISF GC TQI+V LLVGCTECALLAVMSYDRYVAVC
Sbjct:     1 FADLCFSTTTVPQVLVHFLVKRKTISFAGCSTQIVVLLLVGCTECALLAVMSYDRYVAVC 60

Query:   128 KPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCEPPALLKL 187
             KPL+YSTIMT  +C+ L+  SWASGALVSLVDT+FT  LPY G N+INH+FCEPPALLKL
Sbjct:    61 KPLHYSTIMTHWLCVQLAAGSWASGALVSLVDTTFTLRLPYRGNNVINHFFCEPPALLKL 120

Query:   188 ASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFSTCGSHLIV 247
             AS DTYSTEMAIF+MGVVILLAP+SLIL SYWNIISTVIQMQSGEGRLK FSTCGSHLIV
Sbjct:   121 ASADTYSTEMAIFAMGVVILLAPVSLILTSYWNIISTVIQMQSGEGRLKVFSTCGSHLIV 180

Query:   248 VVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTP 283 (SEQ ID NO:142)
             VVLFYGS IF YMRPNSK   E DKMISVFY+AVTP
Sbjct:   181 VVLFYGSAIFAYMRPNSKIMNEKDKMISVFYSAVTP 216 (SEQ ID NO:143)
```

TABLE 117

```
Best hit (human sequence) =

     SWISSPROT-ACC:Q13607 OLFACTORY RECEPTOR-LIKE PROTEIN OLF3 - Homo sapiens
(Human), 317 aa.

                Length = 317

 Score = 840 (295.7 bits), Expect = 9.8e-84, P = 9.8e-84
 Identities = 165/305 (54%), Positives = 217/305 (71%)

Query:     1 MGEENQTFVSKFIFLGLSQDLQTQILLFILFLIIYLLTVLGNQLIIILIFLDSRLHTPMY 60
             MG +NQT+VS+FI LGLS D  T++ LF+LFL++Y++TVLGN LI++LI LDSRLHTPMY
Sbjct:     1 MGTDNQTWVSEFILLGLSSDWDTRVSLFVLFLVMYVVTVLGNCLIVLLIRLDSRLHTPMY 60

Query:    61 FFLRNISFADLCFSTSIVPQVLVHFLVKRKTISFYGCMTQIIVFLLVGCTECALLAVMSY 120
             FFL N+S  D+ ++TS+VPQ+L HFL + K I F  C  Q+   L +G  E  LLAVM+Y
Sbjct:    61 FFLTNLSLVDVSYATSVVPQLLAHFLAEHKAIPFQSCAAQLFFSLALGGIEFVLLAVMAY 120

Query:   121 DRYVAVCKPLYYSTIMTQRVCLWLSFRSWASGALVSLVDTSFTFHLPYWGQNIINHYFCE 180
```

```
              DRYVAVC  L YS IM   +C  L+  SW SG + S V T+ TF LP      I+H  CE
Sbjct:   121 DRYVAVCDALRYSAIMHGGLCARLAITSWVSGFISSPVQTAITFQLPMCRNKFIDHISCE 180

Query:   181 PPALLKLASIDTYSTEMAIFSMGVVILLAPISLILGSYWNIISTVIQMQSGEGRLKAFST 240
             A+++LA +DT S E+ I    +V+L+ P+ L+L SY  IIST++++QS EGR KAF T
Sbjct:   181 LLAVVRLACVDTSSNEVTIMVSSIVLLMTPLCLVLLSYIQIISTILKIQSREGRKKAFHT 240

Query:   241 CGSHLIVVVLFYGSGIFTYMRPNSKTTKELDKMISVFYTAVTPMLNPIIYSLRNKDVKGA 300
             C SHL VV L YG  IFTY++P+S  +   +K+ SVFY  +TPMLNP+IYSLRNK+VKGA
Sbjct:   241 CASHLTVVALCYGVAIFTYIQPHSSPSVLQEKLFSVFYAILTPMLNPMIYSLRNKEVKGA 300

Query:   301 LRKLV 305 (SEQ ID NO:144)
              +KL+
Sbjct:   301 WQKLL 305 (SEQ ID NO:145)
```

Cellular localization analysis using Psort of the protein of invention indicated that it might be targeted to the plasma membrane (certainity =0.6000). The polypeptide seems to have no N-terminal signal sequence. SignalP testing indicated that the most likely cleavage site is between amino acid positions 41 and 42.

SNPs and cSNPs

In the following positions, one or more consensus positions (Cons. Pos.) of the nucleotide sequence have been identified as SNPs (Table 118).

SNPs Position(s):

Possible SNPs found:

200: T->C(2)

127182648(i), phred 36

128123424(i), phred 36

221: A->G(2)

127182619(i), phred 40

128123383(i), phred 29

301: A->G(2)

127182619(i), phred 38

128123383(i), phred 33

330: G->A(2)

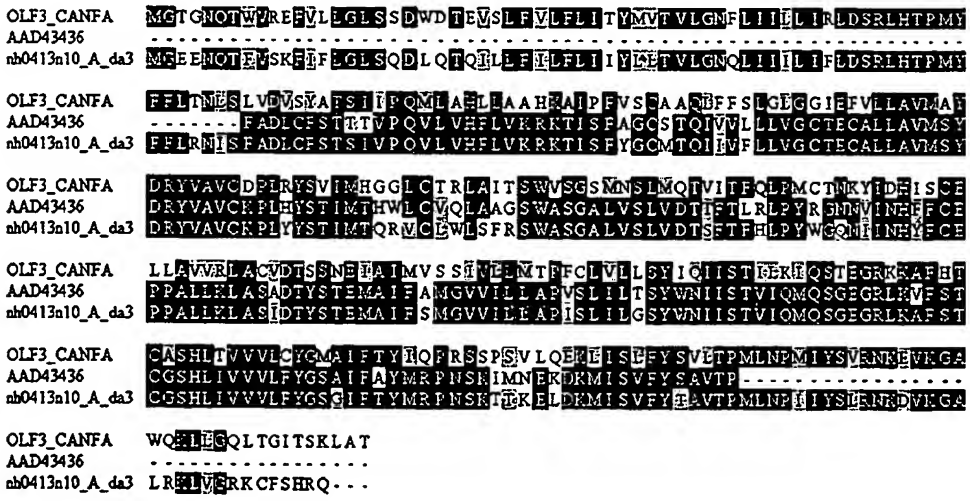
127182648(i), phred 26

128123424(i), phred 26

A multiple sequence alignment is given in Table 118, with the protein of the invention being shown on line 3, in a ClustalW analysis comparing the protein of the invention with related protein sequences. Based on this alignment, black outlined amino acid residues indicate regions of conserved sequence (i.e., regions that may be required to preserve structural or functional properties); greyed amino acid residues can be mutated to a residue with comparable steric and/or chemical properties without altering protein structure or function (e.g. L to V, I, or M); non-highlighted amino acid residues can potentially be mutated to a much broader extent without altering structure or function. Based on its relatedness to members of the GPCR family, the NOV18 protein is a novel member of the OR family. In the following table, nh0413n10_A_da3 is the full-length SEQ ID NO:36. The other sequences in the table are members of the GPCR family and are identified by their Genbank Accession numbers.

TABLE 118

Multiple Alignment:



The nucleic acids and proteins of the invention are useful in potential therapeutic applications implicated in disorders of the neuro-olfactory system, such as those induced by trauma, surgery and/or neoplastic disorders. For example, a cDNA encoding the olfactory receptor protein may be useful in gene therapy for treating such disorders, and the olfactory receptor protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from disorders of the neuro-olfactory system. The novel nucleic acids encoding olfactory receptor protein, and the olfactory receptor protein of the invention, or fragments thereof, may further be useful in the treatment of adenocarcinoma; lymphoma; prostate cancer; uterus cancer, immune response, AIDS, asthma, Crohn's disease, multiple sclerosis, treatment of Albright hereditary ostoeodystrophy, development of powerful assay system for functional analysis of various human disorders which will help in understanding of pathology of the disease, and development of new drug targets for various disorders. They may also be used in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These materials are further useful in the generation of antibodies that bind immunospecifically to the novel substances of the invention for use in therapeutic or diagnostic methods.

NOVX Nucleic Acids

The nucleic acids of the invention include those that encode a NOVX polypeptide or protein. As used herein, the terms polypeptide and protein are interchangeable.

In some embodiments, a NOVX nucleic acid encodes a mature NOVX polypeptide. As used herein, a "mature" form of a polypeptide or protein described herein relates to the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an open reading frame described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an

open reading frame, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

Among the NOVX nucleic acids is the nucleic acid whose sequence is provided in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a fragment thereof. Additionally, the invention includes mutant or variant nucleic acids of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a fragment thereof, any of whose bases may be changed from the corresponding bases shown in Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, while still encoding a protein that maintains at least one of its NOVX-like activities and physiological functions (*i.e.*, modulating angiogenesis, neuronal development). The invention further includes the complement of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, including fragments, derivatives, analogs and homologs thereof. The invention additionally includes nucleic acids or nucleic acid fragments, or complements thereto, whose structures include chemical modifications.

One aspect of the invention pertains to isolated nucleic acid molecules that encode NOVX proteins or biologically active portions thereof. Also included are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNA) and fragments for use as polymerase chain reaction (PCR) primers for the amplification or mutation of NOVX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives,

fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. Examples of isolated nucleic acid molecules include, but are not limited to, recombinant DNA molecules contained in a vector, recombinant DNA molecules maintained in a heterologous host cell, partially or substantially purified nucleic acid molecules, and synthetic DNA or RNA molecules. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecule can contain less than about 50 kb, 25 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a complement of any of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, as a hybridization probe, NOVX nucleic acid sequences can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, eds., MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press,

Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, eds., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at lease 6 contiguous nucleotides of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotide units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der

Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, *e.g.*, a fragment that can be used as a probe or primer, or a fragment encoding a biologically active portion of NOVX. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 98%, or even 99% identity (with a preferred identity of 80-99%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below. An exemplary program is the Gap program (Wisconsin Sequence Analysis Package, Version 8 for UNIX, Genetics Computer Group, University Research Park, Madison, WI) using the default

settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482-489, which is incorporated herein by reference in its entirety).

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of a NOVX polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36, as well as a polypeptide having NOVX activity. Biological activities of the NOVX proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human NOVX polypeptide.

The nucleotide sequence determined from the cloning of the human NOVX gene allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.*, from other tissues, as well as NOVX homologues from other mammals. The probe/primer typically comprises a substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 or more consecutive sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35; or an anti-sense strand nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35; or of a naturally occurring mutant of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35.

Probes based on the human NOVX nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the

probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX gene has been mutated or deleted.

A "polypeptide having a biologically active portion of NOVX" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 that encodes a polypeptide having a NOVX biological activity (biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX. For example, a nucleic acid fragment encoding a biologically active portion of NOVX can optionally include an ATP-binding domain. In another embodiment, a nucleic acid fragment encoding a biologically active portion of NOVX includes one or more regions.

NOVX Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 due to the degeneracy of the genetic code. These nucleic acids thus encode the same NOVX protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 *e.g.*, the polypeptide of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36.

In addition to the human NOVX nucleotide sequence shown in Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, it will be appreciated by those skilled

in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of NOVX may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a NOVX protein, preferably a mammalian NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in NOVX that are the result of natural allelic variation and that do not alter the functional activity of NOVX are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human NOVX cDNA can be isolated based on its homology to human membrane-bound NOVX. Likewise, a membrane-bound human NOVX cDNA can be isolated based on its homology to soluble human NOVX.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500 or 750 nucleotides in length. In another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high

stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions is hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C. This hybridization is followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 corresponds to a naturally occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA 78*: 6789-6792.

Conservative mutations

In addition to naturally-occurring allelic variants of the NOVX sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, thereby leading to changes in the amino acid sequence of the encoded NOVX protein, without altering the functional ability of the NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35. A "non-essential" amino acid residue is a residue that can be altered

from the wild-type sequence of NOVX without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the present invention, are predicted to be particularly unamenable to alteration.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 75% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, or 8. Preferably, the protein encoded by the nucleic acid is at least about 80% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36, more preferably at least about 90%, 95%, 98%, and most preferably at least about 99% homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and

aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in NOVX is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant NOVX protein can be assayed for (1) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically active portions thereof, (2) complex formation between a mutant NOVX protein and a NOVX receptor; (3) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically active portion thereof; (*e.g.*; avidin proteins); (4) the ability to bind NOVX protein; or (5) the ability to specifically bind an anti-NOVX protein antibody.

Antisense NOVX Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "coding region" refers

to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the protein coding region of human NOVX corresponds to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding NOVX. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding NOVX disclosed herein (*e.g.*, Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil,

queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

Such modifications include, by way of nonlimiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

NOVX Ribozymes and PNA moieties

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as a mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide sequence of a NOVX DNA disclosed herein (*i.e.*, Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, NOVX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription of the NOVX gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of NOVX can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*,

DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996); above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl) amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

NOVX Polypeptides

A NOVX polypeptide of the invention includes the NOVX-like protein whose sequence is provided in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residue shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 while still encoding a protein that maintains its NOVX-like activities and physiological functions, or a functional fragment thereof. In some embodiments, up to 20% or more of the residues may be so changed in the mutant or variant protein. In some embodiments, the NOVX polypeptide according to the invention is a mature polypeptide.

In general, a NOVX -like variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an

appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX protein having less than about 30% (by dry weight) of non-NOVX protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX protein, still more preferably less than about 10% of non-NOVX protein, and most preferably less than about 5% non-NOVX protein. When the NOVX protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX protein having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically active portions of a NOVX protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the NOVX protein, *e.g.*, the amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 that include fewer amino acids than the full length NOVX

proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

A biologically active portion of a NOVX protein of the present invention may contain at least one of the above-identified domains conserved between the NOVX proteins, *e.g.* TSR modules. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence shown in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 and retains the functional activity of the protein of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 and retains the functional activity of the NOVX proteins of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36.

Determining homology between two or more sequence

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in either of the sequences being compared for optimal alignment between the sequences). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. See, *Needleman and Wunsch* 1970 *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in Of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region. The term "percentage of positive residues" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical and conservative amino acid substitutions, as defined above, occur in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of positive residues.

Chimeric and fusion proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide operatively linked

to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to NOVX, whereas a "non-NOVX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the NOVX protein and that is derived from the same or a different organism. Within a NOVX fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein. In one embodiment, a NOVX fusion protein comprises at least one biologically active portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at least two biologically active portions of a NOVX protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame to each other. The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX polypeptide.

For example, in one embodiment a NOVX fusion protein comprises a NOVX polypeptide operably linked to the extracellular domain of a second protein. Such fusion proteins can be further utilized in screening assays for compounds that modulate NOVX activity (such assays are described in detail below).

In another embodiment, the fusion protein is a GST-NOVX fusion protein in which the NOVX sequences are fused to the C-terminus of the GST (*i.e.*, glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOVX.

In another embodiment, the fusion protein is a NOVX-immunoglobulin fusion protein in which the NOVX sequences comprising one or more domains are fused to sequences derived from a member of the immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. In one nonlimiting example, a contemplated NOVX ligand of the invention is the NOVX receptor. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e.g.*, cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival, as well as acute and chronic inflammatory disorders and hyperplastic wound healing, e.g. hypertrophic scars and keloids.

Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX agonists and antagonists

The present invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of the NOVX protein. An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form

of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX protein that function as either NOVX agonists (mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of the NOVX protein for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura *et al.* (1984) *Annu Rev Biochem* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucl Acid Res* 11:477.

Polypeptide libraries

In addition, libraries of fragments of the NOVX protein coding sequence can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants (Arkin and Yourvan (1992) PNAS 89:7811-7815; Delgrave *et al.* (1993) Protein Engineering 6:327-331).

NOVX Antibodies

Also included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated NOVX-related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the

amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 2, 4, 6 ,8 ,10, 12, 14, 16, 18, or 20, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX-related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX-related protein sequence will indicate which regions of a NOVX-related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

Polyclonal Antibodies

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen

binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding

specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown *iv vivo* as ascites in a mammal.

The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL

ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as

hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$

fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g.

tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative

mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (Fc R), such as Fc RI (CD64), Fc RII (CD32) and Fc RIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re.

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCL),

active esters (such as disuccinimidyl suberate), aldehydes (such as glutareldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (e.g., avidin) that is in turn conjugated to a cytotoxic agent.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors

(*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or

non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,

SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the

antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as human, Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the

foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

Transgenic NOVX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a

"homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. Sequences including SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the DNA of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35), but more preferably, is a non-human homologue of a human NOVX gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 can be used to construct a homologous recombination vector suitable

for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP

recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The

use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., Proc. Natl. Acad. Sci. USA, 82: 3688 (1985); Hwang et al., Proc. Natl Acad. Sci. USA, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al ., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome. See Gabizon et al., J. National Cancer Inst., 81(19): 1484 (1989).

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use

as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as

pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed. (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York. If the antigenic protein is intracellular and whole antibodies are used as inhibitors, internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used,

the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, *e.g.*, Marasco *et al.*, 1993 *Proc. Natl. Acad. Sci. USA*, 90: 7889-7893. The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended. The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methylmethacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for *iv vivo* administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOT ™ (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (e.g., in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. For example, NOVX activity includes growth and differentiation, antibody production, and tumor growth.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate

substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a NOVX protein or polypeptide of the invention In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the

induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described above.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or

target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also

likely to be involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) identify an individual from a minute biological sample (tissue typing); and (*ii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Tissue Typing

The NOVX sequences of the invention can be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity.

Disorders associated with aberrant NOVX expression of activity include, for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, e.g. Parkinson's disease and Alzheimer's disease.

The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic

DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

One agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies directed against a protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of the protein (e.g., for use in measuring levels of the protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies against the proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antigen binding domain, are utilized as pharmacologically-active compounds.

An antibody specific for a protein of the invention can be used to isolate the protein by standard techniques, such as immunoaffinity chromatography or immunoprecipitation. Such an antibody can facilitate the purification of the natural protein antigen from cells and of recombinantly produced antigen expressed in host cells. Moreover, such an antibody can be used to detect the antigenic protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the antigenic protein. Antibodies directed against the protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*,

physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In one embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The

compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Such disorders include for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, e.g. Parkinson's disease and Alzheimer's disease.

Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from a NOVX gene; (*ii*) an addition of one or more nucleotides to a NOVX gene; (*iii*) a substitution of one or more nucleotides of a NOVX gene, (*iv*) a chromosomal rearrangement of a NOVX gene; (*v*) an alteration in the level of a messenger RNA transcript of a NOVX gene, (*vi*) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (*viii*) a non-wild-type level of a NOVX protein, (*ix*) allelic loss of a NOVX gene, and (*x*) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be

desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in NOVX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the NOVX gene and detect mutations by comparing the sequence of

the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, *e.g.*, a wild-type NOVX sequence,

is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR

amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (*e.g.* disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, e.g. Parkinson's disease and Alzheimer's disease). In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism

of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as

demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting decreased NOVX gene expression, protein levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trails of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other

genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. Disorders associated with aberrant NOVX expression include, for example, disorders of olfactory loss, *e.g.* trauma, HIV illness, neoplastic growth, and neurological disorders, e.g. Parkinson's disease and Alzheimer's disease.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to the subject an agent that modulates NOVX expression or at least one NOVX activity. Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections:-

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates

(*e.g.*, up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable in situations in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated). Another example of such a situation is where the subject has an immunodeficiency disease (*e.g.*, AIDS).

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may used as therapeutic agents. Such agents will generally be employed to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common ranges for

therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1.: Method of Identifying the Nucleic Acids Encoding the G-Protein Coupled Receptors.

The sequence of the invention was derived by laboratory cloning of cDNA fragments covering the full length and/or part of the DNA sequence of the invention, and/or by in silico prediction of the full length and/or part of the DNA sequence of the invention from public human sequence databases. The laboratory cloning was performed using one or more of the methods summarized below:

SeqCallingTM Technology: cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's

proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Exon Linking: The cDNA coding for the sequence was cloned by polymerase chain reaction (PCR) using the following primers: TTACAGAGTTCTGTTTCATTTCCCCTG and AAGATCATTTCCTTTGACTTGTGACCC (SEQ ID NO: 148) on the following pools of human cDNAs: Pool 1 - Adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus.

Primers were designed based on in silico predictions for the full length or part (one or more exons) of the DNA/protein sequence of the invention or by translated homology of the predicted exons to closely related human sequences or to sequences from other species. Usually multiple clones were sequenced to derive the sequence which was then assembled similar to the SeqCalling process. In addition, sequence traces were evaluated manually and edited for corrections if appropriate.

Physical clone: The PCR product derived by exon linking was cloned into the pCR2.1 vector from Invitrogen. The bacterial clone AC019108_E.698009.O17 has an insert covering the entire open reading frame cloned into the pCR2.1 vector from Invitrogen.

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A

SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

Example 2: Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), and Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions).

First, the RNA samples were normalized to constitutively expressed genes such as -actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48°C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using -actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2

min. at 50°C; 10 min. at 95°C; 15 sec. at 95°C/1 min. at 60°C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their -actin /GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions: Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/ l RNase inhibitor, and 0.25 U/ l reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

The plates generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissue were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of

degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

The plates also include samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 µg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) with PHA

(phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^6$ cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol (5.5×10^{-5} M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and +ve selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells

were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^{6} cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 μg/ml or anti-CD40 (Pharmingen) at approximately 10 μg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μg/ml anti-CD28 (Pharmingen) and 2 μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^{5}-10^{6} cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 μg/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 μg/ml) were used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μg/ml) to

prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5 $x10^5$ cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 $x10^5$ cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 µg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 µM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 µl of RNAse-free water and 35 µl buffer (Promega) 5 µl DTT, 7 µl RNAsin and 8 µl DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium

acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C (Table 116)

A. NOV2

The results obtained for NOV2 using the primer-probe set shown below, are shown in Tables 116B, C and D below.

TABLE 116A

Primers	Sequences	TM	Length	Start Position
Forward	5'-TCAAACCTGCTTCTAGCAGTTC-3' (SEQ ID NO:149)	58.8	22	825
Probe	FAM-5'-CTCACCCAGGAGAAAGTATCCTCAGT-3'-TAMRA (SEQ ID NO:150)	64.3	26	848
Reverse	5'-ATCAAGGGATTCAACATGAGAA-3' (SEQ ID NO:151)	58.5	22	891

The results obtained for NOV2 using the above primer-probe set are shown in Tables 116B, C and D below.

TABLE 116B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Liver adenocarcinoma	0.0	40.0	Renal 786-0	0.0	40.0
Heart (fetal)	0.0	40.0	Renal A498	0.0	40.0
Pancreas	100.0	35.1	Renal RXF 393	0.0	40.0
Pancreatic ca. CAPAN 2	0.0	40.0	Renal ACHN	0.0	40.0
Adrenal gland	0.0	40.0	Renal UO-31	0.0	40.0
Thyroid	21.3	37.3	Renal TK-10	0.0	40.0
Salivary gland	0.0	40.0	Liver	0.0	40.0
Pituitary gland	0.0	40.0	Liver (fetal)	0.0	40.0
Brain (fetal)	0.0	40.0	Liver (hepatoblast) HepG2	0.0	40.0
Brain (whole)	0.0	40.0	Lung	0.0	40.0
Brain (amygdala)	8.7	38.6	Lung (fetal)	0.0	40.0
Brain (cerebellum)	0.0	40.0	Lung (small cell) LX-1	0.0	40.0
Brain (hippocampus)	9.9	38.4	Lung (small cell) NCI-H69	0.0	40.0
Brain (thalamus)	0.0	40.0	Lung (s.cell var.) SHP-77	0.0	40.0
Cerebral Cortex	8.6	38.6	Lung (large cell)NCI-H460	0.0	40.0
Spinal cord	0.0	40.0	Lung (non-sm. cell) A549	0.0	40.0
glio/astro U87-MG	14.0	37.9	Lung (non-s.cell) NCI-H23	0.0	40.0
glio/astro U-118-MG	0.0	40.0	Lung (non-s.cell) HOP-62	0.0	40.0
astro SW1783	4.3	39.6	Lung (non-s.cl) NCI-H522	0.0	40.0
neuro; met SK-N-AS	0.0	40.0	Lung (squam.) SW 900	0.0	40.0
astro SF-539	0.0	40.0	Lung (squam.) NCI-H596	0.0	40.0
astro SNB-75	0.0	40.0	Mammary gland	0.0	40.0
glio SNB-19	0.0	40.0	Breast (pl.ef) MCF-7	0.0	40.0
glio U251	0.0	40.0	Breast (pl.ef) MDA-MB-	0.0	40.0

			231		
glio SF-295	0.0	40.0	Breast (pl.ef) T47D	0.0	40.0
Heart	0.0	40.0	Breast BT-549	0.0	40.0
Skeletal muscle	0.0	40.0	Breast MDA-N	0.0	40.0
Bone marrow	0.0	40.0	Ovary	0.0	40.0
Thymus	0.0	40.0	Ovarian OVCAR-3	22.7	37.2
Spleen	0.0	40.0	Ovarian OVCAR-4	0.0	40.0
Lymph node	0.0	40.0	Ovarian OVCAR-5	0.0	40.0
Colorectal	19.5	37.4	Ovarian OVCAR-8	3.8	39.8
Stomach	0.0	40.0	Ovarian IGROV-1	0.0	40.0
Small intestine	0.0	40.0	Ovarian (ascites) SK-OV-3	0.0	40.0
Colon SW480	0.0	40.0	Uterus	0.0	40.0
Colon SW620(SW480 met)	0.0	40.0	Plancenta	19.5	37.4
Colon HT29	17.1	37.6	Prostate	0.0	40.0
Colon HCT-116	0.0	40.0	Prostate (bone met)PC-3	0.0	40.0
Colon CaCo-2	0.0	40.0	Testis	39.2	36.4
Colon Ca tissue(ODO3866)	0.0	40.0	Melanoma Hs688(A).T	0.0	40.0
Colon HCC-2998	0.0	40.0	Melanoma (met) Hs688(B).T	0.0	40.0
Gastric(liver met) NCI-N87	0.0	40.0	Melanoma UACC-62	0.0	40.0
Bladder	20.9	37.3	Melanoma M14	0.0	40.0
Trachea	0.0	40.0	Melanoma LOX IMVI	0.0	40.0
Kidney	0.0	40.0	Melanoma (met) SK-MEL-5	0.0	40.0
Kidney (fetal)	0.0	40.0	Adipose	13.7	37.9

As shown above, NOV2 is expressed in pancreas and testis. Expression levels are quite low (e.g., Ct = 36 for Pancreas).

TABLE 116C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	90.8	33.1	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	13.0	35.9	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	7.4	36.7	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	40.0	Kidney Cancer Clontech 9010320	0.0	40.0
83222 CC NAT (ODO3868)	0.0	40.0	Kidney NAT Clontech 9010321	9.4	36.4
83235 CC Mod Diff	15.9	35.6	Normal Uterus GENPAK	0.0	40.0

(ODO3920)			061018		
83236 CC NAT (ODO3920)	16.8	35.5	Uterus Cancer GENPAK 064011	0.0	40.0
83237 CC Gr.2 ascend colon (ODO3921)	0.0	40.0	Normal Thyroid Clontech A+ 6570-1	0.0	40.0
83238 CC NAT (ODO3921)	85.9	33.2	Thyroid Cancer GENPAK 064010	0.0	40.0
83241 CC from Partial Hepatectomy (ODO4309)	0.0	40.0	Thyroid Cancer INVITROGEN A302152	0.0	40.0
83242 Liver NAT (ODO4309)	0.0	40.0	Thyroid NAT INVITROGEN A302153	0.0	40.0
87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	0.0	40.0
87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0
Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0
84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0	40.0
87073 Prostate Cancer (OD04720-01)	5.1	37.3	GENPAK Breast Cancer 064006	5.8	37.1
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	0.0	40.0	Breast NAT Clontech 9100265	0.0	40.0
83239 Lung Met to Muscle (ODO4286)	0.0	40.0	Breast Cancer INVITROGEN A209073	2.7	38.2
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	2.9	38.1
84136 Lung Malignant Cancer	0.0	40.0	Normal Liver GENPAK	11.0	36.1

(OD03126)			061009		
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	17.6	35.5
84871 Lung Cancer (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0	40.0
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0	40.0
85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	40.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0	40.0
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	100.0	33.0
84139 Melanoma Mets to Lung (OD04321)	0.0	40.0	Bladder Cancer Research Genetics RNA 1023	0.0	40.0
84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	5.3	37.2
Normal Kidney GENPAK 061008	6.2	37.0	87071 Bladder Cancer (OD04718-01)	0.0	40.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	3.8	37.7	87072 Bladder Normal Adjacent (OD04718-03)	0.0	40.0
83787 Kidney NAT (OD04338)	1.3	39.2	Normal Ovary Res. Gen.	0.0	40.0

83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	1.3	39.2
83789 Kidney NAT (OD04339)	0.0	40.0	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	0.0	40.0	Normal Stomach GENPAK 061017	0.0	40.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	0.0	40.0
83793 Kidney NAT (OD04348)	2.2	38.5	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer (OD04622-01)	0.0	40.0	NAT Stomach Clontech 9060394	0.0	40.0
87475 Kidney NAT (OD04622-03)	0.0	40.0	Gastric Cancer Clontech 9060397	0.0	40.0
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	10.4	36.2
85974 Kidney NAT (OD04450-03)	0.0	40.0	Gastric Cancer GENPAK 064005	2.6	38.2
Kidney Cancer Clontech 8120607	0.0	40.0	Kidney NAT Clontech 8120608	0.0	40.0

The results for Table 116C show matches to clone NOV2 in normal bladder, normal colon and two colon cancers as well as colon cancer NAT. Other hits are minor (i.e., have high Ct values). Therefore, NOV2 can be used as a marker in such tissues.

TABLE 116 D

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93099_HUVEC (Endothelial)_starved	0.0	40.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0
93770_Secondary Tr1_anti-	0.0	40.0	93779_HUVEC	0.0	40.0

CD28/anti-CD3			(Endothelial)_IFN gamma		
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93583_Lung Microvascular Endothelial Cells_none	56.6	36.2
93569_primary Th2_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92662_Microvascular Dermal endothelium_none	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	49.0	36.4
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	0.0	40.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	43.5	36.6
93354_CD4_none	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93103_LAK cells_resting	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	0.0	40.0
93788_LAK cells_IL-2	0.0	40.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0
93787_LAK cells_IL-2+IL-12	0.0	40.0	93791_Liver Cirrhosis	48.0	36.5
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93792_Lupus Kidney	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93577_NCI-H292	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0

93578_NK Cells IL-2_resting	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	14.7	38.2	93359_NCI-H292_IL-13	0.0	40.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93357_NCI-H292_IFN gamma	0.0	40.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	0.0	40.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	58.6	36.2
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93249_Ramos (B cell)_none	42.9	36.6	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93250_Ramos (B cell)_ionomycin	0.0	40.0	93256_Normal Human Lung Fibroblast_IL-9	0.0	40.0
93349_B lymphocytes_PWM	0.0	40.0	93255_Normal Human Lung Fibroblast_IL-13	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	0.0	40.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93106_Dermal Fibroblasts CCD1070_resting	0.0	40.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93775_Dendritic Cells_anti-CD40	0.0	40.0	93771_dermal fibroblast_IL-4	0.0	40.0
93774_Monocytes_resting	0.0	40.0	93259_IBD Colitis 1**	0.0	40.0
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93260_IBD Colitis 2	97.9	35.4
93581_Macrophages_resting	0.0	40.0	93261_IBD Crohns	0.0	40.0
93582_Macrophages_LPS 100 ng/ml	0.0	40.0	735010_Colon_normal	0.0	40.0
93098_HUVEC (Endothelial)_none	45.7	36.5	735019_Lung_none	48.0	36.5

The expression levels for the tissues in Table 116D generally are low. The cells that show expression include IBD Colitis 2, HPAEC activated by IL-1 beta and TNF alpha, Liver Cirrhosis, astrocytes activated by TNF alpha and IL1b, and Small Airway Epithelium activated by TNFa and IL1b. Based on the results illustrated in the Tables above, there is potential utility for clone AC019108 C as a marker for colon cancer and in inflammatory conditions (Table 117).

B. NOV3

The results obtained for NOV3 using the primer-probe set shown below, are shown in Tables 117B, C and D below.

TABLE 117A

Primers	Sequences	TM	Length	Start Position
Forward	5'-CACTCTGCTGTCCAGACCATAT-3' (SEQ ID NO:152)	59.2	22	495
Probe	TET-5'-TGACTTTCCATTTGCCCTACTGTGGA-3'-TAMRA (SEQ ID NO:153)	68.6	26	517
Reverse	5'-CGTCACAGAAGTAGTGCTGGAT-3' (SEQ ID NO:154)	59.4	22	552

TABLE 117B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Liver adenocarcinoma	0.0	40.0	Renal 786-0	0.0	40.0
Heart (fetal)	0.0	40.0	Renal A498	0.0	40.0
Pancreas	24.7	33.9	Renal RXF 393	0.0	40.0
Pancreatic ca. CAPAN 2	100.0	31.8	Renal ACHN	0.0	40.0
Adrenal gland	0.0	40.0	Renal UO-31	0.0	40.0
Thyroid	0.0	40.0	Renal TK-10	0.0	40.0
Salivary gland	0.0	40.0	Liver	0.0	40.0
Pituitary gland	0.0	40.0	Liver (fetal)	0.0	40.0
Brain (fetal)	0.0	40.0	Liver (hepatoblast) HepG2	0.0	40.0
Brain (whole)	0.0	40.0	Lung	0.0	40.0
Brain (amygdala)	4.7	36.3	Lung (fetal)	0.0	40.0
Brain (cerebellum)	0.0	40.0	Lung (small cell) LX-1	0.0	40.0
Brain (hippocampus)	0.0	40.0	Lung (small cell) NCI-H69	0.0	40.0
Brain (thalamus)	0.0	40.0	Lung (s.cell var.) SHP-77	0.0	40.0
Cerebral Cortex	2.6	37.1	Lung (large cell)NCI-H460	0.0	40.0
Spinal cord	0.0	40.0	Lung (non-sm. cell) A549	1.4	38.0
glio/astro U87-MG	0.0	40.0	Lung (non-s.cell) NCI-H23	0.0	40.0
glio/astro U-118-MG	1.7	37.7	Lung (non-s.cell) HOP-62	0.0	40.0
astro SW1783	0.0	40.0	Lung (non-s.cl) NCI-H522	0.0	40.0
neuro; met SK-N-AS	0.0	40.0	Lung (squam.) SW 900	0.0	40.0
astro SF-539	0.0	40.0	Lung (squam.) NCI-H596	0.0	40.0
astro SNB-75	0.0	40.0	Mammary gland	0.0	40.0
glio SNB-19	0.0	40.0	Breast (pl.ef) MCF-7	0.0	40.0
glio U251	0.0	40.0	Breast (pl.ef) MDA-MB-231	0.0	40.0
glio SF-295	0.0	40.0	Breast (pl.ef) T47D	2.4	37.2
Heart	0.0	40.0	Breast BT-549	0.0	40.0
Skeletal muscle	0.0	40.0	Breast MDA-N	0.0	40.0
Bone marrow	0.0	40.0	Ovary	0.0	40.0
Thymus	0.0	40.0	Ovarian OVCAR-3	0.0	40.0
Spleen	0.0	40.0	Ovarian OVCAR-4	0.0	40.0
Lymph node	0.0	40.0	Ovarian OVCAR-5	0.0	40.0
Colorectal	19.3	34.2	Ovarian OVCAR-8	0.0	40.0
Stomach	0.0	40.0	Ovarian IGROV-1	0.0	40.0

Small intestine	0.0	40.0	Ovarian (ascites) SK-OV-3	0.0	40.0
Colon SW480	0.0	40.0	Uterus	1.8	37.7
Colon SW620(SW480 met)	0.0	40.0	Plancenta	0.0	40.0
Colon HT29	0.0	40.0	Prostate	0.0	40.0
Colon HCT-116	0.0	40.0	Prostate (bone met)PC-3	0.0	40.0
Colon CaCo-2	0.0	40.0	Testis	0.7	39.1
Colon Ca tissue(ODO3866)	0.0	40.0	Melanoma Hs688(A).T	2.5	37.2
Colon HCC-2998	0.0	40.0	Melanoma (met) Hs688(B).T	0.0	40.0
Gastric(liver met) NCI-N87	0.0	40.0	Melanoma UACC-62	0.0	40.0
Bladder	13.4	34.7	Melanoma M14	0.0	40.0
Trachea	0.0	40.0	Melanoma LOX IMVI	0.0	40.0
Kidney	0.0	40.0	Melanoma (met) SK-MEL-5	2.5	37.2
Kidney (fetal)	0.0	40.0	Adipose	0.0	40.0

As shown in Table 117B above, NOV3 is expressed in a pancreatic carcinoma cell line with high expression (Ct=31); other samples show low or no expression.

TABLE 117C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	47.3	33.1	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	5.7	36.1	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	24.7	34.0	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	40.0	Kidney Cancer Clontech 9010320	0.0	40.0
83222 CC NAT (ODO3868)	5.3	36.2	Kidney NAT Clontech 9010321	0.0	40.0
83235 CC Mod Diff (ODO3920)	5.8	36.1	Normal Uterus GENPAK 061018	0.0	40.0
83236 CC NAT (ODO3920)	23.5	34.1	Uterus Cancer GENPAK 064011	4.1	36.6
83237 CC Gr.2 ascend colon (ODO3921)	1.6	38.0	Normal Thyroid Clontech A+ 6570-1	0.0	40.0
83238 CC NAT (ODO3921)	3.4	36.9	Thyroid Cancer GENPAK 064010	0.0	40.0
83241 CC from Partial Hepatectomy (ODO4309)	4.7	36.4	Thyroid Cancer INVITROGEN A302152	0.0	40.0
83242 Liver NAT (ODO4309)	0.0	40.0	Thyroid NAT INVITROGEN A302153	0.0	40.0

87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	4.2	36.6
87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0
Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0
84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0	40.0
87073 Prostate Cancer (OD04720-01)	3.2	37.0	GENPAK Breast Cancer 064006	0.0	40.0
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	6.8	35.9	Breast NAT Clontech 9100265	2.4	37.4
83239 Lung Met to Muscle (ODO4286)	17.2	34.5	Breast Cancer INVITROGEN A209073	4.9	36.3
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	0.0	40.0
84136 Lung Malignant Cancer (OD03126)	0.0	40.0	Normal Liver GENPAK 061009	4.8	36.4
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	0.0	40.0
84871 Lung Cancer (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	4.3	36.5
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0	40.0
85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	40.0
83255 Ocular Mel Met to Liver (ODO4310)	9.2	35.4	Paired Liver Tissue Research Genetics RNA 6005-N	0.0	40.0
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	100.0	32.0
84139 Melanoma Mets to Lung (OD04321)	4.2	36.6	Bladder Cancer Research Genetics RNA 1023	0.0	40.0

84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	0.0	40.0
Normal Kidney GENPAK 061008	17.4	34.5	87071 Bladder Cancer (OD04718-01)	0.0	40.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	3.8	36.7	87072 Bladder Normal Adjacent (OD04718-03)	0.0	40.0
83787 Kidney NAT (OD04338)	0.0	40.0	Normal Ovary Res. Gen.	0.5	39.6
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	0.0	40.0
83789 Kidney NAT (OD04339)	4.9	36.3	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	0.0	40.0	Normal Stomach GENPAK 061017	0.0	40.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	0.0	40.0
83793 Kidney NAT (OD04348)	8.3	35.6	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer (OD04622-01)	0.0	40.0	NAT Stomach Clontech 9060394	0.0	40.0
87475 Kidney NAT (OD04622-03)	0.0	40.0	Gastric Cancer Clontech 9060397	0.0	40.0
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	7.3	35.8
85974 Kidney NAT (OD04450-03)	0.0	40.0	Gastric Cancer GENPAK 064005	12.4	35.0
Kidney Cancer Clontech 8120607	0.0	40.0			

As shown in Table 117C above, NOV3 is expressed most strongly in normal bladder and clusters to normal colon and colon Cancer. This expression pattern is similar to that for NOV2 above.

TABLE 117D

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0

93769_Secondary Th2_anti-CD28/anti-CD3	0.0	40.0	93779_HUVEC (Endothelial)_IFN gamma	0.0	40.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	40.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	18.2	37.1	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93583_Lung Microvascular Endothelial Cells_none	0.0	40.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	40.0	92662_Microvascular Dermal endothelium_none	0.0	40.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	11.3	37.8	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0

93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	0.0	40.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	11.7	37.8	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0
93354_CD4_none	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	0.0	40.0
93103_LAK cells_resting	0.0	40.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0
93788_LAK cells_IL-2	3.0	39.7	93791_Liver Cirrhosis	100.0	34.7
93787_LAK cells_IL-2+IL-12	12.8	37.6	93792_Lupus Kidney	0.0	40.0
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93577_NCI-H292	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93578_NK Cells IL-2_resting	0.0	40.0	93359_NCI-H292_IL-13	0.0	40.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93357_NCI-H292_IFN gamma	14.8	37.4
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	0.0	40.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	54.7	35.5

93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93249_Ramos (B cell)_none	33.0	36.3	93256_Normal Human Lung Fibroblast_IL-9	0.0	40.0
93250_Ramos (B cell)_ionomycin	0.0	40.0	93255_Normal Human Lung Fibroblast_IL-13	0.0	40.0
93349_B lymphocytes_PWM	3.0	39.7	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	11.0	37.8	93106_Dermal Fibroblasts CCD1070_resting	0.0	40.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	40.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93771_dermal fibroblast_IL-4	0.0	40.0
93775_Dendritic Cells_anti-CD40	0.0	40.0	93259_IBD Colitis 1**	8.7	38.2
93774_Monocytes_resting	0.0	40.0	93260_IBD Colitis 2	34.4	36.2
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93261_IBD Crohns	0.0	40.0
93581_Macrophages_resting	0.0	40.0	735010_Colon_normal	29.9	36.4

93582_Macrophages_LPS 100 ng/ml	16.8	37.2	735019_Lung_none	0.0	40.0
93098_HUVEC (Endothelial)_none	9.0	38.1	64028-1_Thymus_none	14.2	37.5
93099_HUVEC (Endothelial)_starved	0.0	40.0	64030-1_Kidney_none	0.0	40.0

As shown above, there is no discernible pattern of expression. NOV3 is is expressed in cirrhotic liver (low expression, Ct=34.5).

C. NOV4

The results obtained for NOV4 using the primer-probe set shown below, are shown in Tables 118B, C and D below. There is potential utility NOV4 as a diagnostic marker for colon cancer (Table 118)

TABLE 118A

Primers	Sequences	TM	Length	Start Position
Forward	5'-GGATCACACCATCTCATTTGTC-3' (SEQ ID NO:155)	59.3	22	324
Probe	FAM-5'-AGCTGCATCATCCAGTCCTACCTCTA-3'-TAMRA (SEQ ID NO:156)	65.6	26	346
Reverse	5'-GAAGAAGTCAGTGGTGCCTAGA-3' (SEQ ID NO:157)	58.6	22	378

TABLE 118B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Liver adenocarcinoma	0.0	40.0	Renal 786-0	0.0	40.0
Heart (fetal)	0.0	40.0	Renal A498	0.0	40.0
Pancreas	100.0	34.6	Renal RXF 393	0.0	40.0
Pancreatic ca. CAPAN 2	0.0	40.0	Renal ACHN	9.1	38.1
Adrenal gland	9.3	38.0	Renal UO-31	0.0	40.0
Thyroid	0.0	40.0	Renal TK-10	0.0	40.0
Salivary gland	0.0	40.0	Liver	0.0	40.0
Pituitary gland	0.0	40.0	Liver (fetal)	0.0	40.0

Brain (fetal)	0.0	40.0	Liver (hepatoblast) HepG2	0.0	40.0
Brain (whole)	8.5	38.2	Lung	0.0	40.0
Brain (amygdala)	0.0	40.0	Lung (fetal)	0.0	40.0
Brain (cerebellum)	0.0	40.0	Lung (small cell) LX-1	0.0	40.0
Brain (hippocampus)	44.1	35.8	Lung (small cell) NCI-H69	0.0	40.0
Brain (thalamus)	0.0	40.0	Lung (s.cell var.) SHP-77	0.0	40.0
Cerebral Cortex	2.5	39.9	Lung (large cell)NCI-H460	0.0	40.0
Spinal cord	0.0	40.0	Lung (non-sm. cell) A549	7.9	38.3
glio/astro U87-MG	0.0	40.0	Lung (non-s.cell) NCI-H23	0.0	40.0
glio/astro U-118-MG	0.0	40.0	Lung (non-s.cell) HOP-62	0.0	40.0
astro SW1783	0.0	40.0	Lung (non-s.cl) NCI-H522	0.0	40.0
neuro; met SK-N-AS	0.0	40.0	Lung (squam.) SW 900	0.0	40.0
astro SF-539	0.0	40.0	Lung (squam.) NCI-H596	0.0	40.0
astro SNB-75	0.0	40.0	Mammary gland	0.0	40.0
glio SNB-19	0.0	40.0	Breast (pl.ef) MCF-7	0.0	40.0
glio U251	0.0	40.0	Breast (pl.ef) MDA-MB-231	0.0	40.0
glio SF-295	0.0	40.0	Breast (pl.ef) T47D	20.5	36.9
Heart	8.8	38.1	Breast BT-549	0.0	40.0
Skeletal muscle	0.0	40.0	Breast MDA-N	0.0	40.0
Bone marrow	0.0	40.0	Ovary	8.1	38.2
Thymus	0.0	40.0	Ovarian OVCAR-3	0.0	40.0
Spleen	0.0	40.0	Ovarian OVCAR-4	0.0	40.0
Lymph node	0.0	40.0	Ovarian OVCAR-5	0.0	40.0
Colorectal	32.5	36.2	Ovarian OVCAR-8	0.0	40.0
Stomach	0.0	40.0	Ovarian IGROV-1	0.0	40.0

Small intestine	0.0	40.0	Ovarian (ascit s) SK-OV-3	0.0	40.0
Colon SW480	0.0	40.0	Uterus	0.0	40.0
Colon SW620(SW480 met)	0.0	40.0	Plancenta	28.9	36.4
Colon HT29	7.9	38.3	Prostate	0.0	40.0
Colon HCT-116	0.0	40.0	Prostate (bone met)PC-3	0.0	40.0
Colon CaCo-2	7.5	38.3	Testis	32.3	36.2
Colon Ca tissue(ODO3866)	0.0	40.0	Melanoma Hs688(A).T	0.0	40.0
Colon HCC-2998	0.0	40.0	Melanoma (met) Hs688(B).T	0.0	40.0
Gastric(liver met) NCI-N87	0.0	40.0	Melanoma UACC-62	0.0	40.0
Bladder	67.8	35.2	Melanoma M14	0.0	40.0
Trachea	17.4	37.1	Melanoma LOX IMVI	0.0	40.0
Kidney	0.0	40.0	Melanoma (met) SK-MEL-5	0.0	40.0
Kidney (fetal)	0.0	40.0	Adipose	0.0	40.0

As shown above, NOV4 is expressed highest in the pancreas. Other minor (low level) expression is seen in the bladder and colon.

TABLE 118C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	86.5	32.8	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	4.8	36.9	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	8.4	36.1	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid	0.0	40.0	Kidney Cancer Clontech	2.7	37.7

(ODO3868)			9010320		
83222 CC NAT (ODO3868)	2.8	37.7	Kidney NAT Clontech 9010321	2.7	37.8
83235 CC Mod Diff (ODO3920)	9.1	36.0	Normal Uterus GENPAK 061018	0.0	40.0
83236 CC NAT (ODO3920)	52.5	33.5	Uterus Cancer GENPAK 064011	0.0	40.0
83237 CC Gr.2 ascend colon (ODO3921)	1.5	38.6	Normal Thyroid Clontech A+ 6570-1	2.1	38.1
83238 CC NAT (ODO3921)	25.5	34.5	Thyroid Cancer GENPAK 064010	2.7	37.8
83241 CC from Partial Hepatectomy (ODO4309)	0.0	40.0	Thyroid Cancer INVITROGEN A302152	0.0	40.0
83242 Liver NAT (ODO4309)	3.1	37.5	Thyroid NAT INVITROGEN A302153	5.7	36.7
87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	1.1	39.0
87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0
Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0
84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0	40.0
87073 Prostate Cancer (OD04720-01)	0.0	40.0	GENPAK Breast Cancer 064006	0.0	40.0
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	3.2	37.5	Breast NAT Clontech 9100265	0.0	40.0
83239 Lung Met to Muscle	8.4	36.1	Breast Cancer	2.5	37.8

(ODO4286)			INVITROGEN A209073		
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	0.0	40.0
84136 Lung Malignant Cancer (OD03126)	0.0	40.0	Normal Liver GENPAK 061009	0.0	40.0
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	5.8	36.7
84871 Lung Cancer (OD04404)	8.7	36.1	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0	40.0
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0	40.0
85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	40.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0	40.0
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	100.0	32.5
84139 Melanoma Mets to Lung (OD04321)	2.5	37.9	Bladder Cancer Research Genetics RNA 1023	4.1	37.2
84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	0.0	40.0
Normal Kidney GENPAK 061008	3.3	37.5	87071 Bladder Cancer (OD04718-01)	0.0	40.0

83786 Kidney Ca, Nuclear grade 2 (OD04338)	2.4	37.9	87072 Bladder Normal Adjacent (OD04718-03)	0.0	40.0
83787 Kidney NAT (OD04338)	14.3	35.4	Normal Ovary Res. Gen.	0.0	40.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	2.8	37.7
83789 Kidney NAT (OD04339)	2.1	38.2	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	2.5	37.8	Normal Stomach GENPAK 061017	1.7	38.4
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	0.0	40.0
83793 Kidney NAT (OD04348)	4.5	37.0	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer (OD04622-01)	2.4	38.0	NAT Stomach Clontech 9060394	2.9	37.6
87475 Kidney NAT (OD04622-03)	0.0	40.0	Gastric Cancer Clontech 9060397	0.0	40.0
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	4.5	37.0
85974 Kidney NAT (OD04450-03)	6.6	36.5	Gastric Cancer GENPAK 064005	10.7	35.8
Kidney Cancer Clontech 8120607	0.0	40.0			

As shown above, NOV4 is expressed strongly in bladder and clusters to normal colon and two colon cancer samples (higher expression in NAT colon). Consistent, but low, expression is seen in kidney NAT margin as well (Cts 37-38).

TABLE 118D

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	40.0	93779_HUVEC (Endothelial)_IFN gamma	0.0	40.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	40.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	40.0	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93583_Lung Microvascular Endothelial Cells_none	0.0	40.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	40.0	92662_Microvascular Dermal endothelium_none	0.0	40.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0

93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	40.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	43.2	35.5
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	17.1	36.8
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	20.5	36.6
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0
93354_CD4_none	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	0.0	40.0
93103_LAK cells_resting	7.5	38.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0
93788_LAK cells_IL-2	0.0	40.0	93791_Liver Cirrhosis	4.2	38.9
93787_LAK cells_IL-2+IL-12	0.0	40.0	93792_Lupus Kidney	28.3	36.1
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93577_NCI-H292	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93578_NK Cells IL-2_resting	0.0	40.0	93359_NCI-H292_IL-13	0.0	40.0
93109_Mixed Lymphocyte	0.0	40.0	93357_NCI-H292_IFN	0.0	40.0

Reaction_Two Way MLR			gamma		
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	50.0	35.3
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	100.0	34.3
93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93249_Ramos (B cell)_none	0.0	40.0	93256_Normal Human Lung Fibroblast_IL-9	10.1	37.6
93250_Ramos (B cell)_ionomycin	0.0	40.0	93255_Normal Human Lung Fibroblast_IL-13	10.4	37.5
93349_B lymphocytes_PWM	7.1	38.1	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	7.5	38.0	93106_Dermal Fibroblasts CCD1070_resting	0.0	40.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	40.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93771_dermal fibroblast_IL-4	0.0	40.0

93775_Dendritic Cells_anti-CD40	0.0	40.0	93259_IBD Colitis 1**	0.0	40.0
93774_Monocytes_resting	0.0	40.0	93260_IBD Colitis 2	24.5	36.3
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93261_IBD Crohns	0.0	40.0
93581_Macrophages_resting	0.0	40.0	735010_Colon_normal	0.0	40.0
93582_Macrophages_LPS 100 ng/ml	0.0	40.0	735019_Lung_none	0.0	40.0
93098_HUVEC (Endothelial)_none	0.0	40.0	64028-1_Thymus_none	0.0	40.0
93099_HUVEC (Endothelial)_starved	0.0	40.0	64030-1_Kidney_none	0.0	40.0

As shown above, NOV4 is expressed in human pulmonary arterial endothelial cells resting and stimulated with IL-1b and TNF-a (Ct=34; fairly low level).

D. NOV5

The results obtained for NOV5 using the primer-probe set shown below, are shown in Tables 119B, C and D below. There is potential utility of NOV5 as a marker for colon cancer and in therapy of colon cancer (Table 119).

TABLE 119A

Primers	Sequences	TM	Length	Start Position
Forward	5'-AAGAAATGAAGGCAGCATTACA-3' (SEQ ID NO:158)	58.9	22	1000
Probe	TET-5'-CACAAGGAAGTGCAGCCTCACTGACT-3'-TAMRA (SEQ ID NO: 159)	69.1	26	1035
Reverse	5'-CATTATAGGGTTTCCTGCATGT-3' (SEQ ID NO:160)	58	22	1074

TABLE 119B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Liver adenocarcinoma	0.0	40.0	Renal 786-0	0.0	40.0
Heart (fetal)	0.0	40.0	Renal A498	0.0	40.0

Pancreas	88.9	35.4	Renal RXF 393	0.0	40.0
Pancreatic ca. CAPAN 2	0.0	40.0	Renal ACHN	0.0	40.0
Adrenal gland	0.0	40.0	Renal UO-31	0.0	40.0
Thyroid	0.0	40.0	Renal TK-10	0.0	40.0
Salivary gland	0.0	40.0	Liver	0.0	40.0
Pituitary gland	0.0	40.0	Liver (fetal)	14.3	38.0
Brain (fetal)	0.0	40.0	Liver (hepatoblast) HepG2	0.0	40.0
Brain (whole)	0.0	40.0	Lung	0.0	40.0
Brain (amygdala)	0.0	40.0	Lung (fetal)	0.0	40.0
Brain (cerebellum)	0.0	40.0	Lung (small cell) LX-1	0.0	40.0
Brain (hippocampus)	0.0	40.0	Lung (small cell) NCI-H69	0.0	40.0
Brain (thalamus)	0.0	40.0	Lung (s.cell var.) SHP-77	0.0	40.0
Cerebral Cortex	0.0	40.0	Lung (large cell)NCI-H460	0.0	40.0
Spinal cord	0.0	40.0	Lung (non-sm. cell) A549	0.0	40.0
glio/astro U87-MG	0.0	40.0	Lung (non-s.cell) NCI-H23	0.0	40.0
glio/astro U-118-MG	0.0	40.0	Lung (non-s.cell) HOP-62	0.0	40.0
astro SW1783	0.0	40.0	Lung (non-s.cl) NCI-H522	0.0	40.0
neuro; met SK-N-AS	0.0	40.0	Lung (squam.) SW 900	0.0	40.0
astro SF-539	0.0	40.0	Lung (squam.) NCI-H596	100.0	35.2
astro SNB-75	0.0	40.0	Mammary gland	0.0	40.0
glio SNB-19	0.0	40.0	Breast (pl.ef) MCF-7	0.0	40.0
glio U251	0.0	40.0	Breast (pl.ef) MDA-MB-231	0.0	40.0
glio SF-295	16.3	37.8	Breast (pl.ef) T47D	0.0	40.0
Heart	0.0	40.0	Breast BT-549	0.0	40.0
Skeletal muscle	0.0	40.0	Breast MDA-N	0.0	40.0

Bone marrow	0.0	40.0	Ovary	0.0	40.0
Thymus	0.0	40.0	Ovarian OVCAR-3	0.0	40.0
Spleen	0.0	40.0	Ovarian OVCAR-4	0.0	40.0
Lymph node	0.0	40.0	Ovarian OVCAR-5	0.0	40.0
Colorectal	55.9	36.0	Ovarian OVCAR-8	0.0	40.0
Stomach	0.0	40.0	Ovarian IGROV-1	0.0	40.0
Small intestine	0.0	40.0	Ovarian (ascites) SK-OV-3	0.0	40.0
Colon SW480	0.0	40.0	Uterus	0.0	40.0
Colon SW620(SW480 met)	0.0	40.0	Plancenta	0.0	40.0
Colon HT29	0.0	40.0	Prostate	0.0	40.0
Colon HCT-116	0.0	40.0	Prostate (bone met)PC-3	0.0	40.0
Colon CaCo-2	0.0	40.0	Testis	64.2	35.8
Colon Ca tissue(ODO3866)	0.0	40.0	Melanoma Hs688(A).T	11.3	38.3
Colon HCC-2998	0.0	40.0	Melanoma (met) Hs688(B).T	0.0	40.0
Gastric(liver met) NCI-N87	0.0	40.0	Melanoma UACC-62	0.0	40.0
Bladder	43.8	36.4	Melanoma M14	0.0	40.0
Trachea	0.0	40.0	Melanoma LOX IMVI	0.0	40.0
Kidney	0.0	40.0	Melanoma (met) SK-MEL-5	0.0	40.0
Kidney (fetal)	0.0	40.0	Adipose	0.0	40.0

As shown above, NOV5 shows highest expression in a squamous cell carcinoma of the lung. In addition it is expressed in testis, pancreas, colon and bladder.

TABLE 119C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	74.2	34.1	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	0.0	40.0	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	18.7	36.1	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	40.0	Kidney Cancer Clontech 9010320	0.0	40.0
83222 CC NAT (ODO3868)	0.0	40.0	Kidney NAT Clontech 9010321	0.0	40.0
83235 CC Mod Diff (ODO3920)	0.0	40.0	Normal Uterus GENPAK 061018	0.0	40.0
83236 CC NAT (ODO3920)	6.8	37.6	Uterus Cancer GENPAK 064011	0.0	40.0
83237 CC Gr.2 ascend colon (ODO3921)	8.4	37.2	Normal Thyroid Clontech A+ 6570-1	12.5	36.7
83238 CC NAT (ODO3921)	16.5	36.3	Thyroid Cancer GENPAK 064010	0.0	40.0
83241 CC from Partial Hepatectomy (ODO4309)	0.0	40.0	Thyroid Cancer INVITROGEN A302152	10.7	36.9
83242 Liver NAT (ODO4309)	0.0	40.0	Thyroid NAT INVITROGEN A302153	9.4	37.1
87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	17.6	36.2
87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0
Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0

84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	2.2	39.2
87073 Prostate Cancer (OD04720-01)	0.0	40.0	GENPAK Breast Cancer 064006	0.0	40.0
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	0.0	40.0	Breast NAT Clontech 9100265	0.0	40.0
83239 Lung Met to Muscle (ODO4286)	7.1	37.5	Breast Cancer INVITROGEN A209073	0.0	40.0
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	0.0	40.0
84136 Lung Malignant Cancer (OD03126)	0.0	40.0	Normal Liver GENPAK 061009	7.9	37.3
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	0.0	40.0
84871 Lung Cancer (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	17.3	36.2
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	7.6	37.4
85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	40.0
83255 Ocular Mel Met to Liver	0.0	40.0	Paired Liver Tissue Research Genetics RNA	0.0	40.0

(ODO4310)			6005-N		
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	100.0	33.7
84139 Melanoma Mets to Lung (OD04321)	0.0	40.0	Bladder Cancer Research Genetics RNA 1023	0.0	40.0
84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	0.0	40.0
Normal Kidney GENPAK 061008	5.5	37.9	87071 Bladder Cancer (OD04718-01)	0.0	40.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	4.4	38.2	87072 Bladder Normal Adjacent (OD04718-03)	8.9	37.2
83787 Kidney NAT (OD04338)	12.9	36.6	Normal Ovary Res. Gen.	0.0	40.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	0.0	40.0
83789 Kidney NAT (OD04339)	13.7	36.5	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	0.0	40.0	Normal Stomach GENPAK 061017	0.0	40.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	0.0	40.0
83793 Kidney NAT (OD04348)	0.0	40.0	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer (OD04622-01)	0.0	40.0	NAT Stomach Clontech 9060394	0.0	40.0
87475 Kidney NAT (OD04622-03)	0.0	40.0	Gastric Cancer Clontech 9060397	0.0	40.0
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	0.0	40.0

85974 Kidney NAT (OD04450-03)	15.8	36.3	Gastric Cancer GENPAK 064005	0.0	40.0
Kidney Cancer Clontech 8120607	0.0	40.0			

As shown above, NOV5 shows highest expression in normal bladder (Ct=33.7) and normal colon (Ct=34.1). There are low levels of expression (Ct>36) in colon cancers and colon cancer NAT, normal and cancerous thyroid and two kidney NAT samples.

TABLE 119D

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	40.0	93779_HUVEC (Endothelial)_IFN gamma	0.0	40.0
93770_Secondary Tr1_anti-CD28/anti-CD3	17.0	37.9	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	40.0	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93583_Lung Microvascular Endothelial Cells_none	0.0	40.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93569_primary Th2_anti-CD28/anti-	0.0	40.0	92662_Microvascular Dermal	0.0	40.0

CD3			endothelium_none		
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	40.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	19.6	37.7
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0
93354_CD4_none	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	0.0	40.0
93103_LAK cells_resting	0.0	40.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0

93788_LAK cells_IL-2	0.0	40.0	93791_Liver Cirrhosis	56.6	36.2
93787_LAK cells_IL-2+IL-12	0.0	40.0	93792_Lupus Kidney	0.0	40.0
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93577_NCI-H292	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93578_NK Cells IL-2_resting	0.0	40.0	93359_NCI-H292_IL-13	0.0	40.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93357_NCI-H292_IFN gamma	0.0	40.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	0.0	40.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	100.0	35.4
93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93249_Ramos (B cell)_none	0.0	40.0	93256_Normal Human Lung Fibroblast_IL-9	0.0	40.0
93250_Ramos (B cell)_ionomycin	0.0	40.0	93255_Normal Human Lung Fibroblast_IL-13	0.0	40.0
93349_B lymphocytes_PWM	0.0	40.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	0.0	40.0	93106_Dermal Fibroblasts CCD1070_resting	0.0	40.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93361_Dermal Fibroblasts	0.0	40.0

			CCD1070_TNF alpha 4 ng/ml		
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93771_dermal fibroblast_IL-4	0.0	40.0
93775_Dendritic Cells_anti-CD40	0.0	40.0	93259_IBD Colitis 1**	18.1	37.8
93774_Monocytes_resting	0.0	40.0	93260_IBD Colitis 2	18.3	37.8
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93261_IBD Crohns	0.0	40.0
93581_Macrophages_resting	0.0	40.0	735010_Colon_normal	0.0	40.0
93582_Macrophages_LPS 100 ng/ml	0.0	40.0	735019_Lung_none	0.0	40.0
93098_HUVEC (Endothelial)_none	0.0	40.0	64028-1_Thymus_none	15.9	38.0
93099_HUVEC (Endothelial)_starved	0.0	40.0	64030-1_Kidney_none	0.0	40.0

As shown above, NOV5 is expressed in human pulmonary arterial endothelial cells stimulated with IL-1b and TNF-a (Ct=35.4; fairly low level).

E. NOV6

The results obtained for NOV6 using the primer-probe set shown below, are shown in Tables 120B, C and D below. There is potential utility of NOV6 as a marker for liver cancer and in therapy of liver cancer (Table 120).

TABLE 120A

Primers	Sequences	TM	Length	Start Position
Forward	5'-CCTCCACACCACCATGTACTAC-3'(SEQ ID NO: 161)	59.3	22	183
Probe	FAM-5'-TCACCAACCTGTCGTTCATTGACATG-3'-TAMRA (SEQ ID NO:162)	69.1	26	209
R verse	5'-GAACACCAAAGTCATCAGCAAT-3' (SEQ ID NO:163)	59.1	22	261

TABLE 120B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	16.7	38.1	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	17.6	38.0	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	9.1	39.0	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	40.0	Kidney Cancer Clontech 9010320	0.0	40.0
83222 CC NAT (ODO3868)	18.8	37.9	Kidney NAT Clontech 9010321	0.0	40.0
83235 CC Mod Diff (ODO3920)	8.3	39.1	Normal Uterus GENPAK 061018	0.0	40.0
83236 CC NAT (ODO3920)	0.0	40.0	Uterus Cancer GENPAK 064011	0.0	40.0
83237 CC Gr.2 ascend colon (ODO3921)	25.7	37.5	Normal Thyroid Clontech A+ 6570-1	0.0	40.0
83238 CC NAT (ODO3921)	15.4	38.2	Thyroid Cancer GENPAK 064010	0.0	40.0
83241 CC from Partial Hepatectomy (ODO4309)	0.0	40.0	Thyroid Cancer INVITROGEN A302152	0.0	40.0
83242 Liver NAT (ODO4309)	0.0	40.0	Thyroid NAT INVITROGEN A302153	0.0	40.0
87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	0.0	40.0
87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0

Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0
84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0	40.0
87073 Prostate Cancer (OD04720-01)	0.0	40.0	GENPAK Breast Cancer 064006	0.0	40.0
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	0.0	40.0	Breast NAT Clontech 9100265	0.0	40.0
83239 Lung Met to Muscle (ODO4286)	10.5	38.8	Breast Cancer INVITROGEN A209073	13.8	38.4
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	0.0	40.0
84136 Lung Malignant Cancer (OD03126)	0.0	40.0	Normal Liver GENPAK 061009	0.0	40.0
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	7.0	39.4
84871 Lung Cancer (OD04404)	9.2	39.0	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0	40.0
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0	40.0
85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research	0.0	40.0

			Genetics RNA 6005-T		
83255 Ocular Mel Met to Liver (ODO4310)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0	40.0
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	100.0	35.5
84139 Melanoma Mets to Lung (OD04321)	0.0	40.0	Bladder Cancer Research Genetics RNA 1023	0.0	40.0
84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	8.8	39.0
Normal Kidney GENPAK 061008	6.2	39.5	87071 Bladder Cancer (OD04718-01)	15.8	38.2
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.0	40.0	87072 Bladder Normal Adjacent (OD04718-03)	0.0	40.0
83787 Kidney NAT (OD04338)	0.0	40.0	Normal Ovary Res. Gen.	0.0	40.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	0.0	40.0
83789 Kidney NAT (OD04339)	9.0	39.0	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	0.0	40.0	Normal Stomach GENPAK 061017	0.0	40.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	0.0	40.0
83793 Kidney NAT (OD04348)	0.0	40.0	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer (OD04622-01)	0.0	40.0	NAT Stomach Clontech 9060394	0.0	40.0
87475 Kidney NAT	0.0	40.0	Gastric Cancer Clontech	0.0	40.0

(OD04622-03)			9060397		
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	0.0	40.0
85974 Kidney NAT (OD04450-03)	12.7	38.5	Gastric Cancer GENPAK 064005	0.0	40.0
Kidney Cancer Clontech 8120607	0.0	40.0			

As shown above, NOV6 is expressed to a high level in normal bladder. It also clusters to colon cancers and colon NAT, with generally low levels of expression (Ct=37-38).

TABLE 120C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	40.0	93779_HUVEC (Endothelial)_IFN gamma	0.0	40.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	40.0	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	40.0	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93583_Lung Microvascular Endothelial Cells_none	0.0	40.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and	0.0	40.0

			IL1b (1 ng/ml)		
93569_primary Th2_anti-CD28/anti-CD3	0.0	40.0	92662_Microvascular Dermal endothelium_none	0.0	40.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	40.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	0.0	40.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0
93354_CD4_none	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	0.0	40.0

93103_LAK cells_resting	0.0	40.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0
93788_LAK cells_IL-2	0.0	40.0	93791_Liver Cirrhosis	100.0	36.3
93787_LAK cells_IL-2+IL-12	0.0	40.0	93792_Lupus Kidney	0.0	40.0
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93577_NCI-H292	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93578_NK Cells IL-2_resting	0.0	40.0	93359_NCI-H292_IL-13	0.0	40.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93357_NCI-H292_IFN gamma	0.0	40.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	0.0	40.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	0.0	40.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93249_Ramos (B cell)_none	0.0	40.0	93256_Normal Human Lung Fibroblast_IL-9	0.0	40.0
93250_Ramos (B cell)_ionomycin	0.0	40.0	93255_Normal Human Lung Fibroblast_IL-13	0.0	40.0
93349_B lymphocytes_PWM	0.0	40.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	38.7	37.7	93106_Dermal Fibroblasts	0.0	40.0

			CCD1070_resting		
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	40.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93771_dermal fibroblast_IL-4	0.0	40.0
93775_Dendritic Cells_anti-CD40	0.0	40.0	93259_IBD Colitis 1**	0.0	40.0
93774_Monocytes_resting	0.0	40.0	93260_IBD Colitis 2	0.0	40.0
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93261_IBD Crohns	0.0	40.0
93581_Macrophages_resting	0.0	40.0	735010_Colon_normal	0.0	40.0
93582_Macrophages_LPS 100 ng/ml	0.0	40.0	735019_Lung_none	0.0	40.0
93098_HUVEC (Endothelial)_none	0.0	40.0	64028-1_Thymus_none	12.8	39.3
93099_HUVEC (Endothelial)_starved	0.0	40.0	64030-1_Kidney_none	0.0	40.0

As shown above, The highest level of expression of NOV6 (at a low level; Ct=36.3) occurs in cirrhotic liver.

F. NOV10

The results obtained for NOV10 using the primer-probe set shown below, are shown in Tables 121B, C and D below. There is potential utility of NOV10 as a marker for colon cancer and in therapy of colon cancer (Table 121)

TABLE 121A

Primers	Sequences	TM	Length	Start Position
Forward	5'-ACTACGTGCCACCTGTCTGTAT-3'(SEQ ID NO:164)	58.7	22	832

Probe	TET-5'-CTACCTGCAGCCTCGCTCCAGTGAG-3'-TAMRA (SEQ ID NO:165)	71.4	25	854
Reverse	5'-AGCATTGGAGTTACGATTGTGT-3' (SEQ ID NO:166)	58.7	22	907

TABLE 121B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Normal Colon GENPAK 061003	21.6	36.5	Kidney NAT Clontech 8120608	0.0	40.0
83219 CC Well to Mod Diff (ODO3866)	10.2	37.6	Kidney Cancer Clontech 8120613	0.0	40.0
83220 CC NAT (ODO3866)	19.2	36.6	Kidney NAT Clontech 8120614	0.0	40.0
83221 CC Gr.2 rectosigmoid (ODO3868)	0.0	40.0	Kidney Cancer Clontech 9010320	0.0	40.0
83222 CC NAT (ODO3868)	0.0	40.0	Kidney NAT Clontech 9010321	0.0	40.0
83235 CC Mod Diff (ODO3920)	12.9	37.2	Normal Uterus GENPAK 061018	0.0	40.0
83236 CC NAT (ODO3920)	10.4	37.5	Uterus Cancer GENPAK 064011	0.0	40.0
83237 CC Gr.2 ascend colon (ODO3921)	0.0	40.0	Normal Thyroid Clontech A+ 6570-1	0.0	40.0
83238 CC NAT (ODO3921)	0.0	40.0	Thyroid Cancer GENPAK 064010	0.0	40.0
83241 CC from Partial Hepatectomy (ODO4309)	0.0	40.0	Thyroid Cancer INVITROGEN A302152	0.0	40.0
83242 Liver NAT (ODO4309)	0.0	40.0	Thyroid NAT INVITROGEN A302153	0.0	40.0
87472 Colon mets to lung (OD04451-01)	0.0	40.0	Normal Breast GENPAK 061019	0.0	40.0

87473 Lung NAT (OD04451-02)	0.0	40.0	84877 Breast Cancer (OD04566)	0.0	40.0
Normal Prostate Clontech A+ 6546-1	0.0	40.0	85975 Breast Cancer (OD04590-01)	0.0	40.0
84140 Prostate Cancer (OD04410)	0.0	40.0	85976 Breast Cancer Mets (OD04590-03)	0.0	40.0
84141 Prostate NAT (OD04410)	0.0	40.0	87070 Breast Cancer Metastasis (OD04655-05)	0.0	40.0
87073 Prostate Cancer (OD04720-01)	0.0	40.0	GENPAK Breast Cancer 064006	0.0	40.0
87074 Prostate NAT (OD04720-02)	0.0	40.0	Breast Cancer Clontech 9100266	0.0	40.0
Normal Lung GENPAK 061010	14.2	37.1	Breast NAT Clontech 9100265	0.0	40.0
83239 Lung Met to Muscle (ODO4286)	0.0	40.0	Breast Cancer INVITROGEN A209073	0.0	40.0
83240 Muscle NAT (ODO4286)	0.0	40.0	Breast NAT INVITROGEN A2090734	0.0	40.0
84136 Lung Malignant Cancer (OD03126)	0.0	40.0	Normal Liver GENPAK 061009	0.0	40.0
84137 Lung NAT (OD03126)	0.0	40.0	Liver Cancer GENPAK 064003	0.0	40.0
84871 Lung Cancer (OD04404)	100.0	34.3	Liver Cancer Research Genetics RNA 1025	0.0	40.0
84872 Lung NAT (OD04404)	0.0	40.0	Liver Cancer Research Genetics RNA 1026	0.0	40.0
84875 Lung Cancer (OD04565)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0	40.0
85950 Lung Cancer (OD04237-01)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0	40.0

85970 Lung NAT (OD04237-02)	0.0	40.0	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	40.0
83255 Ocular Mel Met to Liver (ODO4310)	0.0	40.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.0	40.0
83256 Liver NAT (ODO4310)	0.0	40.0	Normal Bladder GENPAK 061001	43.2	35.5
84139 Melanoma Mets to Lung (OD04321)	0.0	40.0	Bladder Cancer Research Genetics RNA 1023	0.0	40.0
84138 Lung NAT (OD04321)	0.0	40.0	Bladder Cancer INVITROGEN A302173	0.0	40.0
Normal Kidney GENPAK 061008	8.8	37.8	87071 Bladder Cancer (OD04718-01)	0.0	40.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.0	40.0	87072 Bladder Normal Adjacent (OD04718-03)	13.2	37.2
83787 Kidney NAT (OD04338)	0.0	40.0	Normal Ovary Res. Gen.	0.0	40.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	40.0	Ovarian Cancer GENPAK 064008	0.0	40.0
83789 Kidney NAT (OD04339)	0.0	40.0	87492 Ovary Cancer (OD04768-07)	0.0	40.0
83790 Kidney Ca, Clear cell type (OD04340)	0.0	40.0	87493 Ovary NAT (OD04768-08)	0.0	40.0
83791 Kidney NAT (OD04340)	14.2	37.1	Normal Stomach GENPAK 061017	0.0	40.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	40.0	NAT Stomach Clontech 9060359	11.3	37.4
83793 Kidney NAT (OD04348)	0.0	40.0	Gastric Cancer Clontech 9060395	0.0	40.0
87474 Kidney Cancer	0.0	40.0	NAT Stomach Clontech	0.0	40.0

(OD04622-01)			9060394		
87475 Kidney NAT (OD04622-03)	0.0	40.0	Gastric Cancer Clontech 9060397	0.0	40.0
85973 Kidney Cancer (OD04450-01)	0.0	40.0	NAT Stomach Clontech 9060396	12.4	37.3
85974 Kidney NAT (OD04450-03)	2.3	39.7	Gastric Cancer GENPAK 064005	0.0	40.0
Kidney Cancer Clontech 8120607	0.0	40.0			

As shown above, the highest expression of NOV10 is in one lung cancer with no detectable expression in the cognate NAT, and normal bladder (Ct=35.5). There is a low level cluster to colon cancers and colon NAT.

TABLE 121C

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	40.0	93100_HUVEC (Endothelial)_IL-1b	0.0	40.0
93769_Secondary Th2_anti-CD28/anti-CD3	11.3	38.8	93779_HUVEC (Endothelial)_IFN gamma	0.0	40.0
93770_Secondary Tr1_anti-CD28/anti-CD3	5.4	39.8	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	40.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	40.0	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	40.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	40.0	93781_HUVEC (Endothelial)_IL-11	0.0	40.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	40.0	93583_Lung Microvascular Endothelial	0.0	40.0

			Cells_none		
93568_primary Th1_anti-CD28/anti-CD3	0.0	40.0	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	40.0	92662_Microvascular Dermal endothelium_none	0.0	40.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	40.0	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	40.0	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	0.0	40.0
93566_primary Th2_resting dy 4-6 in IL-2	0.0	40.0	93347_Small Airway Epithelium_none	0.0	40.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	40.0	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	100.0	35.6
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92668_Coronery Artery SMC_resting	16.7	38.2
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.0	40.0	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	40.0	93107_astrocytes_resting	0.0	40.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	40.0	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	40.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	40.0	92666_KU-812 (Basophil)_resting	0.0	40.0

93354_CD4_none	0.0	40.0	92667_KU-812 (Basophil)_PMA/ionoycin	0.0	40.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	40.0	93579_CCD1106 (Keratinocytes)_none	14.3	38.4
93103_LAK cells_resting	0.0	40.0	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	40.0
93788_LAK cells_IL-2	0.0	40.0	93791_Liver Cirrhosis	65.5	36.2
93787_LAK cells_IL-2+IL-12	0.0	40.0	93792_Lupus Kidney	0.0	40.0
93789_LAK cells_IL-2+IFN gamma	0.0	40.0	93577_NCI-H292	0.0	40.0
93790_LAK cells_IL-2+ IL-18	0.0	40.0	93358_NCI-H292_IL-4	0.0	40.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	40.0	93360_NCI-H292_IL-9	0.0	40.0
93578_NK Cells IL-2_resting	0.0	40.0	93359_NCI-H292_IL-13	0.0	40.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93357_NCI-H292_IFN gamma	0.0	40.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93777_HPAEC_-	0.0	40.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	40.0	93778_HPAEC_IL-1 beta/TNA alpha	0.0	40.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	40.0	93254_Normal Human Lung Fibroblast_none	0.0	40.0
93113_Mononuclear Cells (PBMCs)_PWM	0.0	40.0	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	40.0
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	40.0	93257_Normal Human Lung Fibroblast_IL-4	0.0	40.0
93249_Ramos (B cell)_none	0.0	40.0	93256_Normal Human Lung Fibroblast_IL-9	0.0	40.0
93250_Ramos (B cell)_ionomycin	17.6	38.1	93255_Normal Human Lung Fibroblast_IL-13	0.0	40.0

93349_B lymphocytes_PWM	0.0	40.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.0	40.0
93350_B lymphoytes_CD40L and IL-4	17.3	38.2	93106_Dermal Fibroblasts CCD1070_resting	0.0	40.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	40.0	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	40.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	40.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	40.0
93356_Dendritic Cells_none	0.0	40.0	93772_dermal fibroblast_IFN gamma	0.0	40.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	40.0	93771_dermal fibroblast_IL-4	24.5	37.7
93775_Dendritic Cells_anti-CD40	0.0	40.0	93259_IBD Colitis 1**	0.0	40.0
93774_Monocytes_resting	0.0	40.0	93260_IBD Colitis 2	0.0	40.0
93776_Monocytes_LPS 50 ng/ml	0.0	40.0	93261_IBD Crohns	0.0	40.0
93581_Macrophages_resting	0.0	40.0	735010_Colon_normal	0.0	40.0
93582_Macrophages_LPS 100 ng/ml	0.0	40.0	735019_Lung_none	0.0	40.0
93098_HUVEC (Endothelial)_none	0.0	40.0	64028-1_Thymus_none	0.0	40.0
93099_HUVEC (Endothelial)_starved	0.0	40.0	64030-1_Kidney_none	0.0	40.0

As shown above, the highest expression for NOV10 is found in IL-1 and TNF-a stimulated small airway epithelium (Ct=35.6) and cirrhotic liver (Ct=36.2).

G. NOV12

The results obtained for NOV12 using the primer-probe set shown below, are shown in Tables 122B, C and D below. There is potential utility of NOV12 as a marker for skin and/or breast cancer and in therapy of skin and/or breast cancer (Table 122)

TABLE 122A

Primers	Sequences	Length	Start Position
Forward	5'-CTGTCTGCAAGCCCCTGTACTAC-3' (SEQ ID NO:161)	23	404
Probe	FAM-5'-TGGCCCAGCTGACCCTGCTCA-3'-TAMRA (SEQ ID NO:162)	21	
Reverse	5'-GGCCCAGGACCTGAAGGA-3' (SEQ ID NO:163)	18	463

TABLE 122B

Tissue_Name	Rel. Expr.%	Ct value	Tissue_Name	Rel. Expr.%	Ct value
Endothelial cells	0.0	40.0	Kidney (fetal)	0.0	40.0
Endothelial cells (treated)	0.0	40.0	Renal ca.786-0	0.0	40.0
Pancreas	0.0	40.0	Renal ca.A498	0.0	40.0
Pancreatic ca.CAPAN 2	0.0	40.0	Renal ca. RXF 393	0.0	40.0
Adipose	100.0	28.0	Renal ca. ACHN	0.0	40.0
Adrenal gland	0.0	40.0	Renal ca. UO-31	0.0	40.0
Thyroid	0.0	40.0	Renal ca. TK-10	0.0	40.0
Salavary gland	0.0	40.0	Liver	0.0	40.0
Pituitary gland	0.0	40.0	Liver (fetal)	0.0	40.0
Brain (fetal)	0.0	40.0	Liver ca. (hepatoblast) HepG2	0.0	40.0
Brain (whole)	0.0	40.0	Lung	0.0	40.0
Brain (amygdala)	0.0	40.0	Lung (fetal)	0.0	40.0
Brain (cerebellum)	0.0	40.0	Lung ca. (small cell) LX-1	0.0	40.0

Brain (hippocampus)	0.0	40.0	Lung ca. (small cell) NCI-H69	0.0	40.0
Brain (substantia nigra)	0.0	40.0	Lung ca. (s.cell var.) SHP-77	0.0	40.0
Brain (thalamus)	0.0	40.0	Lung ca. (large cell)NCI-H460	0.0	40.0
Brain (hypothalamus)	0.0	40.0	Lung ca. (non-sm. cell) A549	0.0	40.0
Spinal cord	0.0	40.0	Lung ca. (non-s.cell) NCI-H23	0.0	40.0
CNS ca. (glio/astro) U87-MG	0.0	40.0	Lung ca (non-s.cell) HOP-62	0.0	40.0
CNS ca. (glio/astro) U-118-MG	0.0	40.0	Lung ca. (non-s.cl) NCI-H522	0.0	40.0
CNS ca. (astro)SW1783	0.0	40.0	Lung ca. (squam.)SW 900	0.0	40.0
CNS ca.* (neuro; met) SK-N-AS	0.0	40.0	Lung ca. (squam.) NCI-H596	0.0	40.0
CNS ca. (astro) SF-539	0.0	40.0	Mammary gland	0.0	40.0
CNS ca. (astro) SNB-75	0.0	40.0	Breast ca.* (pl. effusion) MCF-7	0.0	40.0
CNS ca. (glio)SNB-19	0.0	40.0	Breast ca.* (pl.ef) MDA-MB-231	0.0	40.0
CNS ca. (glio)U251	0.0	40.0	Breast ca.* (pl. effusion) T47D	0.0	40.0
CNS ca. (glio)SF-295	0.0	40.0	Breast ca. BT-549	0.0	40.0
Heart	0.0	40.0	Breast ca. MDA-N	3.9	32.7
Skeletal muscle	0.0	40.0	Ovary	0.0	40.0
Bone marrow	0.0	40.0	Ovarian ca.OVCAR-3	0.0	40.0
Thymus	0.0	40.0	Ovarian ca.OVCAR-4	0.0	40.0
Spleen	0.0	40.0	Ovarian ca.OVCAR-5	0.0	40.0

Lymph node	0.0	40.0	Ovarian ca. OVCAR-8	0.0	40.0
Colon (ascending)	14.1	30.9	Ovarian ca. IGROV-1	0.0	40.0
Stomach	0.0	40.0	Ovarian ca.* (ascites) SK-OV-3	0.0	40.0
Small intestine	0.0	40.0	Uterus	0.0	40.0
Colon ca. SW480	0.0	40.0	Plancenta	0.0	40.0
Colon ca.* (SW480 met)SW620	0.0	40.0	Prostate	0.0	40.0
Colon ca. HT29	0.0	40.0	Prostate ca.* (bone met)PC-3	0.0	40.0
Colon ca.HCT-116	0.0	40.0	Testis	0.0	40.0
Colon ca. CaCo-2	0.0	40.0	Melanoma Hs688(A).T	0.0	40.0
Colon ca. HCT-15	0.0	40.0	Melanoma* (met) Hs688(B).T	0.0	40.0
Colon ca. HCC-2998	0.1	37.7	Melanoma UACC-62	0.0	40.0
Gastric ca.* (liver met) NCI-N87	0.0	40.0	Melanoma M14	0.0	40.0
Bladder	0.0	40.0	Melanoma LOX IMVI	0.0	40.0
Trachea	0.0	40.0	Melanoma* (met) SK-MEL-5	0.0	40.0
Kidney	0.0	40.0	Melanoma SK-MEL-28	16.8	30.6

As shown above, the highest expression of NOV12 occurs in adipose, but genomic DNA contamination is suspected. Also, significant expression is found in melanoma cell line SK-MEL-28, one breast cancer cell line and in ascending colon.

OTHER EMBODIMENTS

While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
   a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36;
   b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;
   c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36;
   d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and
   e) a fragment of any of a) through d).

2. The polypeptide of claim 1 that is a naturally occurring allelic variant of the sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36.

3. The polypeptide of claim 2, wherein the variant is the translation of a single nucleotide polymorphism.

4. The polypeptide of claim 1 that is a variant polypeptide described therein, wherein any amino acid specified in the chosen sequence is changed to provide a conservative substitution.

5. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

a) a mature form of the amino acid sequence given SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36;

b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed;

c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36;

d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed;

e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and

f) the complement of any of said nucleic acid molecules.

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5 that encodes a variant polypeptide, wherein the variant polypeptide has the polypeptide sequence of a naturally occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a single nucleotide polymorphism encoding said variant polypeptide.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of

a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35;

b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed;

c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35; and

d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

10. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 and/or 35, or a complement of said nucleotide sequence.

11. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence in which any nucleotide specified in the coding sequence of the chosen nucleotide sequence is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides in the chosen coding

sequence are so changed, an isolated second polynucleotide that is a complement of the first polynucleotide, or a fragment of any of them.

12. A vector comprising the nucleic acid molecule of claim 11.

13. The vector of claim 12, further comprising a promoter operably linked to said nucleic acid molecule.

14. A cell comprising the vector of claim 12.

15. An antibody that binds immunospecifically to the polypeptide of claim 1.

16. The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17. The antibody of claim 15, wherein the antibody is a humanized antibody.

18. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing said sample;

(b) introducing said sample to a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

20. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

(a) introducing said polypeptide to said agent; and

(b) determining whether said agent binds to said polypeptide.

21. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

(a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;

(b) contacting the cell with a composition comprising a candidate substance; and

(c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic agent.

22. A method for modulating the activity of the polypeptide of claim 1, the method comprising introducing a cell sample expressing the polypeptide of said claim with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

23. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent said pathology in said subject.

24. The method of claim 23, wherein said subject is a human.

25. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering to a subject in which such treatment or prevention is desired a NOVX nucleic acid in an amount sufficient to treat or prevent said pathology in said subject.

26. The method of claim 25, wherein said subject is a human.

27. A method of treating or preventing a pathology associated with the polypeptide of claim 1, said method comprising administering to a subject in which such treatment or prevention is desired a NOVX antibody in an amount sufficient to treat or prevent said pathology in said subject.

28. The method of claim 27, wherein the subject is a human.

29. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically acceptable carrier.

30. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically acceptable carrier.

31. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically acceptable carrier.

32. A kit comprising in one or more containers, the pharmaceutical composition of claim 29.

33. A kit comprising in one or more containers, the pharmaceutical composition of claim 30.

34. A kit comprising in one or more containers, the pharmaceutical composition of claim 31.

35. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is the polypeptide of claim 1.

36. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is a NOVX nucleic acid.

37. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein said therapeutic is a NOVX antibody.

38. A method for screening for a modulator of activity or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;

b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and

c) comparing the activity of said protein in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator of latency of, or predisposition to, a pathology associated with the polypeptide of claim 1.

39. The method of claim 38, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

40. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

b) comparing the amount of said polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

41. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 5 in a first mammalian subject, the method comprising:

a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and

b) comparing the amount of said nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

42. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 and/or 36 or a biologically active fragment thereof.

43. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the pathological state.